Metaphorical Polysemy of the Chinese Color Term hēi “black”
A corpus-based cognitive semantic analysis with Behavioral Profiles
Liu, Meichun; Dou, Jinmeng

Published in:
International Journal of Corpus Linguistics

Published: 01/02/2024

Document Version:
Post-print, also known as Accepted Author Manuscript, Peer-reviewed or Author Final version

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.1075/ijcl.21067.liu

Publication details:

Citing this paper
Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights
Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission
Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.
Metaphorical Polysemy of the Chinese Color Term  hēi  黑 “black”

A corpus-based cognitive semantic analysis with Behavioral Profiles

Meichun Liu¹, Jinmeng Dou²

¹ meichliu@cityu.edu.hk, Department of Linguistics and Translation, City University of Hong Kong, Hong Kong, China

² Corresponding author: jmdou2-c@my.cityu.edu.hk, Department of Linguistics and Translation, City University of Hong Kong, Hong Kong, China

This paper provides a corpus-based, cognitive semantic study on profiling the varied uses of the Chinese color term  hēi  黑 “black” with regard to its ‘metaphorical polysemy’. We hypothesize that the semantic (dis)similarities among the eight metaphorical meanings of  hēi  “black” can be captured by clustering their contextual features, including collocational patterns, morphosyntactic and semantic properties, and discourse information. The ‘Behavioral Profiles’ approach is adopted for the analyses with the annotations of 800 instances for 46 contextual features and a ‘hierarchical agglomerative cluster analysis’ conducted on the annotated data. The result shows that the eight metaphorical senses of  hēi  “black” fall into three clusters. The clustering can be explained by the conceptual bases pertaining to color perceptions and color changes, in line with the ‘Conceptual Metaphor Theory’. This study demonstrates the effectiveness of the corpus-based BP approach in exploring the underlying cognitive mechanisms of metaphorical extensions and meaning differentiations.
1. Introduction

It is well recognized that color perceptions provide common sources of metaphors. Color terms (CTs) may refer to emotional states, political stance, or other semantic domains beyond their literal meanings referring to natural colors in perception or physics. More noteworthy is that the use of CTs in a given context may render more than one metaphorical meaning, which is viewed as ‘metaphorical polysemy’ (Apresjan, 1974; Jurafsky, 1996). Chinese CTs are no exception. For instance, the meaning of the Chinese CT  hēi 黑 “black”, as one of the earliest-acquired CTs in most languages (Berlin & Kays, 1969; Wu, 2011), continues to be extended over time through metaphorical extension. The following example provides an initial illustration of metaphorical polysemy of  hēi “black”, which may be interpreted with its literal meaning of the black color, or two other metaphorical meanings:

(1) 这位运动员实在是太黑了
zhè-wèi yùndòngyuán shì tài tài hēi le
this-CL athlete really SHI too black LE
(“Lit.: This athlete is truly too black.”) or
(“Met.: This athlete is too evil or too outstanding.”)

In this example, the syntactic structure is quite clear: the term  hēi “black” is used with a degree adverb to predicate the subject zhè-wèi yùndòngyuán “this athlete”. However, out of context, the exact meaning of this sentence may be unclear and ambiguous, since there are two possible metaphorical readings of  hēi “black” in addition to its literal meaning, as shown below in Table 1.
Table 1. Possible interpretations of hēi “black” in example (1)

<table>
<thead>
<tr>
<th>Possible Meaning</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black skin color</td>
<td>the literal meaning of hēi 黑 “black”</td>
</tr>
<tr>
<td>Immoral/evil</td>
<td>derived from hēi-xīn 黑心 “black heart” (metaphor)</td>
</tr>
<tr>
<td>Exceptional/surprising</td>
<td>derived from hēi-mǎ 黑马 “black horse” (metaphor &amp; metonymy)</td>
</tr>
</tbody>
</table>

Besides the two metaphorical meanings in the above example, the CT hēi “black” can also be used for other metaphorical meanings as specified in dictionaries and previous studies. Given the above observation, this paper aims to explore the semantic (dis)similarities among all the possible metaphorical meanings of hēi “black”. A corpus-based Behavioral Profiles (BP) approach is adopted to provide a usage-based, quantitative analysis of semantic profiling on the varied uses of hēi “black” with regard to its metaphorical polysemy. To do so, eight metaphorical senses of hēi “black” are postulated (as shown in Table 2 below) by means of reviewing the meanings proposed in related studies and the Contemporary Chinese Dictionary (7th ed.), and subsequently verified by examining corpus data from Sketch Engine, as explained in Section 3.1.

Following previous BP studies on lexical semantics (Gries, 2006; Divjark & Gries, 2006; Gries & Divjark, 2009), we hypothesize that there must be some underlying semantic relations between the identified metaphorical meanings of hēi “black”. These semantic relations can be captured by the local contexts they are embedded in. Based on ‘Distributional Semantics’ (Harris, 1954; Firth, 1951), 46 contextual features (‘ID tags’), as detailed in Section 3.2, are proposed with taking into consideration of both formal and functional distinctions. They include collocational patterns, morphosyntactic and semantic features, as well as discourse information in actual usages. The BP approach is adopted to conduct a hierarchical agglomerative cluster analysis on the eight proposed metaphorical meanings of hēi “black”, on the basis of a dataset consisting of 800 instances (100 for each meaning) that are annotated with the proposed features.

The paper is organized as follows: After the introduction in Section 1, Section 2 reviews the previous analyses of the metaphorical meanings of hēi “black” and the
BP works on semantic distinctions of polysemy or near-synonyms. Section 3 explicates our research methodology, including the data collection and the annotation of the relevant contextual features. Section 4 presents the results of the cluster analyses on the eight metaphorical meanings of *hēi* “black” and provides cognitive semantic accounts for the statistically significant distinctive features for each cluster.

It is concluded that the underlying semantic relations between the eight metaphorical meanings of *hēi* “black” can be well captured by means of Behavioral Profiles, which demonstrates the effectiveness of the corpus-based BP approach in exploring the cognitive mechanisms of metaphorical extensions.

### 2. Metaphorical Polysemy and Behavioral Profiles

This section firstly reviews the previous studies on metaphorical polysemy of CTs. Then, we provide a review of the linguistic studies on metaphorical polysemy with special attention to the Chinese CT *hēi* “black”. In the third sub-section, an overview of BP studies in cognitive linguistics is provided.

#### 2.1 Metaphorical Polysemy of Color Terms

Polysemy is a universal phenomenon in language, as most lexical items may be polysemous through extensions of their core meanings (Copestake & Briscoe, 1995; Jackendoff, 2002; Murphy, 2002; Pustejovsky, 1995). In the account of polysemy, metaphorical transfer is frequently seen as an underlying mechanism in many previous studies. Besides the extension from the literal to the metaphorical meanings, some lexical items may have multiple metaphorical mappings, resulting in metaphorical polysemy. Apresjan (1974) proposed that metaphorical polysemy is one type of polysemy. The author assumed that there is an analogical relation between the different senses of polysemy. Jurafsky (1996) applied metaphor as one of the tools to structure
the different meanings of the diminutive in modeling the synchronic semantic relations of polysemy within a RADIAL CATEGORY (Lakoff, 1987). Additionally, Klepousniotou (2002) examined the psychological reality of ambiguous words with metaphorical extensions. To explore the diachronic evolution of metaphorical meanings of polysemy, Xu et al. (2017) demonstrated that metaphorical mappings are systemic with measurable communicative and cognitive principles based on a large-scale survey of historical evidence.

Concerning the lexical sources of metaphorical polysemy, CTs constitute a major group. A number of studies were dedicated to exploring the possible metaphorical meanings of CTs with a qualitative approach. They found that some basic CTs do have more than one metaphorical meaning in various languages, such as English (Allan, 2009), European languages (Hill, 2008), Persian (Amouzadeha et al., 2011; Aliakbari & Khosravian, 2012). It was also argued that the metaphorical meanings of CTs can vary from language to language (Wierzbicka, 1990; Ghafel & Mirzaie, 2014; Chatti 2016; Hastürkoğlu, 2018; Al-Jarf, 2019). In other words, although CTs tend to be metaphorically polysemous, some of the metaphorical meanings may be language-specific or culturally exclusive.

Chinese has several basic CTs which are monosyllabic and monomorphemic, such as ǹēi “black”, ǹái “white”, ǹóg “red”, ǹuáng “yellow”, ǹù “green”, etc. As in many other languages, these terms are often used to refer to color-associated evaluations in social beliefs and political ideology. For example, ǹēi “black” is associated with darkness and illegality; and ǹái “white” is associated with light and righteousness. Putting the two terms together, the compound ǹēi-ǹái “black and white” is coined as a cover term for social norms and justice. The red color, ǹóng “red”, can refer to danger or auspicious blessings. The yellow color, ǹuáng “yellow”, represents the royal color or the color for pornographic matters. It is clear that most CTs are polysemous in nature and metaphorically extended to diverse domains.

This paper, as the first study of a series of research studies on metaphorical polysemy of Chinese CTs, focuses on one of the basic CTs, ǹēi “black” in Chinese, with
an aim to explore its potential range of metaphorical uses. The reasons that we choose *hēi* “black” as the research target in this study are threefold.³ Firstly, as mentioned above, *hēi* “black” is one of the earliest-acquired CTs in Chinese (Berlin & Kay, 1969; Wu, 2011). Secondly, the meaning of *hēi* “black” continues to be extended over time through metaphorical extensions as discussed by Zhang (1988), Xing (2008), Li & Bai (2013) and Lai & Chung (2018). Thirdly, the morphosyntactic behaviors and other contextual variations of *hēi* “black” with different metaphorical meanings are quite diverse and flexible. For example, *hēi* “black” can be an attributive modifier (hēi chūzū “black[Illegal] cabs”), a stative predicate (tā liǎn hěn hēi “His face is very black[Be angry]”), a transitive predicate (diànnǎo bèi hēi “The computer was blackened[hacked]”), or lexicalized in a Verb-Resultative compound (lā-hēi “blacklist”). In the next section, we further review the previous studies on the specific Chinese CT, *hēi* “black”, with a focus on the proposed metaphorical meanings of that term.

### 2.2 Metaphorical Polysemy of Chinese Color Term *hēi* “black”

According to the pioneering work of Berlin & Kays (1969), BLACK is one of the earliest-acquired CTs in a language as long as there is a word referring to a color concept in that language. The Chinese CT *hēi* “black” is indeed among the most studied CTs in the literature. As one of the earliest studies on Chinese CTs, Wu (1986) mentioned the polysemous nature of Chinese CTs, focusing on their various symbolic and transferred meanings. Zhang (1988) indicated that many Chinese CTs possess more than one associative meaning from a cultural perspective, which is taken to be an inherent characteristic of CTs. With the development of ‘Conceptual Metaphor Theory’ (Lakoff & Johnson, 1980; Lakoff, 1993), more recent studies analyzed and reinterpreted the extended meanings of Chinese CTs as conceptual metaphors from the perspective of cognitive semantics (Xing, 2008; Li & Bai, 2013; Lai & Chung, 2018). Focusing on the Chinese basic CTs, Wu (2011) demonstrated with diachronic data that the Chinese
character hēi “black” is indeed regarded as a basic CT from Late Shang Dynasty to Modern China (1500 BC-present). With regard to its polysemy, Xing (2008) discussed the mechanisms in motivating its wide range of extended meanings, and indicated that hēi “black” has five additional senses besides its original meaning of the black color. These senses include one extended meaning (Dark) and four abstract meanings (Bad, Malevolent, Secret, and Illegal). In a comparative study of English and Chinese, Li & Bai (2013) listed ten metaphorical meanings of hēi “black”, including Evil, Loss of Consciousness, Reaction, Slander, Illegal, Scar, Negative, Grievance, Angry, and Bad. In another study, Lai & Chung (2018) suggested eleven metaphorical meanings of hēi “black” in Taiwanese Mandarin, including Evil/Vicious, Illegal/Underground, Secret/Mysterious, Low/Sluggish, Disgraceful/Dishonorable, Depressed/Hopeless, Negative/Unfortunate, Keeping a low profile, Sarcastic/Cynical, Unexpectedly excellent and Darkness.

As we can see from the above, a number of metaphorical meanings of hēi “black” have been proposed. However, most previous studies only concern the possible number or range of its extended meanings and regard each meaning as a discrete semantic prime, without considering the potential semantic relations among these distinguished senses. Furthermore, with the rise of social media, some novel metaphorical usages of hēi “black” are under-researched, such as hēi “black” referring to Network Attack (e.g., wǎngzhàn bèi hēi “The website is hacked.”). Just as each CT covers a gradable region in the color spectrum, the different metaphorical meanings of CTs may also be gradable and interrelated with overlaps, which makes it difficult to isolate the senses on pure semantic grounds. It is therefore important to look for lexical-grammatical evidence to differentiate the meanings. To explore the interrelationship among the metaphorical meanings of hēi “black”, this study adopts the BP approach to measure the semantic (dis)similarities between these different meanings quantitatively.

2.3 Behavioral Profiles Studies on Polysemy
The corpus-based BP approach offers an empirical way to study the subtle semantic differences of polysemy, antonymy, or synonymy, under the framework of ‘Distributional Semantics’ (Harris, 1954; Firth, 1951). It incorporates the lexical-grammatical behaviors associated with lexical meanings into the statistic model of feature analysis, which represents the probabilistic tendency of their occurrences in real contexts. This approach is composed of two key notions: ‘ID tags’ and ‘Behavioral Profiles’. ID tags, first proposed by Atkins (1987), refer to the collocation or colligation that correlates with a particular word. Gries (2006) further extended the scope of ID tags to cover more linguistic parameters related to a particular word at different levels, including morphological, syntactic, semantic, and collocational features. The ‘Behavioral Profiles’ provide empirically significant information with various statistical methods. It was firstly proposed by Hanks (1996) as the complementation patterns and roles played by a word. The notion was further extended to an “inventory of elements co-occurring with a word within the confines of a simple clause or sentence in actual speech and writing” by Gries & Divjak (2009: 277).

Further elaborated by Divjak (2003) and Gries (2006), the BP approach has shown high effectiveness on lexical-semantic issues with a series of studies on various pairs or sets of lexical-semantic relations. A frequent implementation of this method is to analyze the (dis)similarities of two or more synonymous items based on their grammatical behaviors (cf. Divjak, 2003, 2010; Liu, 2010; Levshina, 2011; Dosedlová & Lu, 2019; Liesenfeld et al., 2020). Compared to the numerous works on synonymy, fewer BP studies have focused on polysemy, and much less on metaphorical polysemy, which may involve a more complicated model and account. Two such studies were conducted by Gries (2006, 2017), in which two hierarchical cluster analyses were performed on 48 meanings of run and 26 senses of get respectively. In addition, Jansegers et al. (2015) investigated the polysemy (30 meanings) of the Spanish perception verb sentir “feel”. With regard to the dynamic process of semantic change, there are also some diachronic studies on the polysemy of verbs with varied statistical methods in the BP approach. Among them, Ioannou (2020) focused on the semantic
evolutions of two Greek verbs, kāmnō “do” and eutheìazō “make”, with Multidimensional Scaling (MDS) maps. Adopting the same statistic model, Jansegers & Gries (2017) further investigated the diachronic semantic changes of the Spanish perception verb, sentir “feel”. All these works detected statistically significant clustering effects of the metaphorical meanings among the interpretable clusters.

In sum, to our best knowledge, there has not been a study yet that demonstrates the applicability of the BP approach on exploring the underlying cognitive mechanisms of different metaphorical extensions of polysemy.

3. Data and Method

Methodologically, this study regarded each metaphorical meaning of hēi “black” as a gradable variable. Their semantic relations were measured and visualized based on the mutual semantic distance, which were calculated from the annotated corpus data. The raw data contained instances that were extracted from two corpora: ‘The Corpus of Chinese Simplified Web 2017 Sample’ and ‘the Chinese GigaWord 2 Corpus (Mainland, simplified)’. They were accessed through the Sketch Engine platform. The BP approach was then applied to the analysis. Specifically, the BP approach involved several steps, as briefly introduced below.

First, a dataset was created by compiling a corpus of sampled target sentences that contained the lexical item under study, i.e., the morpheme hēi “black”, with different metaphorical meanings. Then, a set of categorical variables, that can characterize the local context of the target lexical item, were defined as ID tags from different levels, such as lexical, sentence, and discourse levels. After annotating the collected sentences with defined variables, each metaphorical meaning was then represented as a vector that comprised of the proportions of the variables ranging from 0 to 1, called BP vectors. Finally, the cluster analysis was conducted to investigate the semantic relations between the proposed metaphorical meanings. Please note that all
the calculations in this paper were conducted with R language.

3.1 Data Collection

To identify the metaphorical meanings of hēi “black”, we first conducted a thorough review of the metaphorical meanings of the term proposed in the previously relevant studies, along with a consultation of ‘the Contemporary Chinese Dictionary (7th ed.)’\(^2\). To empirically verify the previously proposed or enlisted meanings, two corpora, ‘the Corpus of Chinese Simplified Web 2017 Sample’ and ‘the Chinese Gigaword 2 Corpus (Mainland, simplified)’, were then used to examine whether the meanings could be commonly found in the database of actual usages. Note that the former is mainly composed of internet texts and the latter newswires, which provide balanced coverage for the usages of that term. In this way, EIGHT metaphorical meanings of hēi “black” were identified for further analysis in this study. They were shown in Table 2 below with examples selected from the corpora.

<table>
<thead>
<tr>
<th>No.</th>
<th>Meaning &amp; Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Slander, Entrap</td>
</tr>
<tr>
<td></td>
<td>如今,即使是胡歌这样的娱乐圈模范,都被黑了。</td>
</tr>
<tr>
<td></td>
<td>rújīn, jíshì shì Húgē zhè-yàng-de yúlè-quān mófàn, dōu bèi hēi le</td>
</tr>
<tr>
<td></td>
<td>nowadays even SHI NAME this-kind-DE entertainment circle model, still BEI black LE</td>
</tr>
<tr>
<td></td>
<td>“Nowadays, in the entertainment circle, even models like Hu Ge have been slandered.”</td>
</tr>
<tr>
<td>2</td>
<td>Illegal, Underground</td>
</tr>
<tr>
<td></td>
<td>魏善庄村一民房内存在一面包黑作坊。</td>
</tr>
<tr>
<td></td>
<td>wèishànhuāng-cūn yī mín-fáng nèi cúnzài yī miànbāo hēi-zuòfāng</td>
</tr>
<tr>
<td></td>
<td>Weishanzhuang Village one private-house inside exist one bread black workshop</td>
</tr>
<tr>
<td></td>
<td>“There is an illegal bakery in a private house in Weishanzhuang Village.”</td>
</tr>
<tr>
<td>3</td>
<td>Evil, Malevolent</td>
</tr>
<tr>
<td></td>
<td>这里的人心太黑！</td>
</tr>
<tr>
<td></td>
<td>zhèlǐ-de rén-xīn tài hēi</td>
</tr>
<tr>
<td></td>
<td>here-DE human-heart too black</td>
</tr>
<tr>
<td></td>
<td>“The hearts of the people here are too malevolent.”</td>
</tr>
<tr>
<td>No.</td>
<td>Meaning</td>
</tr>
<tr>
<td>-----</td>
<td>------------------</td>
</tr>
<tr>
<td>4</td>
<td>Unfavorable, Bad</td>
</tr>
<tr>
<td>5</td>
<td>Angry, Sullen</td>
</tr>
<tr>
<td>6</td>
<td>Secret, Mysterious</td>
</tr>
<tr>
<td>7</td>
<td>Unexpected, Surprising</td>
</tr>
<tr>
<td>8</td>
<td>Network Attack</td>
</tr>
</tbody>
</table>

Precisely, a total of 23,860 instances containing 黑“black” were extracted from the two corpora as our original database: 17,331 instances were from ‘Chinese Simplified Web 2017 Sample’ and 6,529 instances were from ‘Chinese Gigaword 2 Corpus (Mainland, simplified)’. We manually checked all examples and extracted the instances referring to the eight proposed metaphorical meanings of 黑“black”. Eight sub-datasets, corresponding to the eight metaphorical meanings, were obtained. For each meaning, 100 instances were randomly collected from the sub-dataset, which gave rise to a final dataset of 800 instances. To avoid unnecessary complications, idiomatic and low-frequency expressions were cleaned from the dataset, such as 贝黑-guō “take the
blame for others” and yī-tiáo lù zǒu-dào hēi “stick to one way until dark (to do something)”. Lastly, the 800 collected instances were annotated manually based on the selected contextual features (ID tags). The following section provides an overview of these features.

3.2 Introduction to Contextual Features (ID tags)

As mentioned above, contextual features (ID tags) are one of the essential notions in the BP approach. In line with previous studies (Divjak and Gries, 2006; Gires and Divjak, 2009; Liesenfeld et al., 2020), we proposed a wide range of heterogeneous features with two steps. Firstly, Word Sketches (Kilgarriff & Tugwell, 2001) was used to obtain a preliminary insight of the collocational patterns of hēi “black”. Then, we further enriched the ID tags beyond collocational associations to include other salient features pertaining to the part of speech of hēi “black” and its grammatical role in a sentence. It was observed that the collocational patterns and other lexical-semantic features of hēi “black” are quite varied when referring to different metaphorical meanings. Below is an example to illustrate the annotation:

(2) 沃尔沃卡车如何玩转黑科技?

_Wòērwò kāchē rúhé wán-zhuǎn hēi-kējì_

Volvo truck how play-around black-technology

(“How do Volvo trucks play around with the black technology?”)

In this instance, we first annotated the part of speech of hēi “black” as an attributive adjective, and the syntactic role of hēi as a modifier in the compound NP hēi-kējì “black technology”. Then, the syntactic role of the whole NP was annotated as the object of the sentence. Note that the semantic type and the discourse information of the noun contextually collocated with hēi “black” were also annotated as part of our BP design.

As shown in Table 3 below, a total of 46 contextual features, containing 132
contextual variable levels, fall into three categories: lexical-collocational patterns (25), morphosyntactic and semantic features (14), and discourse information (7). For the lexical-collocational patterns, the selected features pertain to the collocational patterns of *hēi* “black” with other lexical categories, such as degree or negation markers. The morphosyntactic and semantic features concern mainly the syntactic categories, the grammatical functions, and the semantic types of the surrounding elements in the same constituent, including the part of speech of *hēi* “black”, the syntactic roles of *hēi* “black”, the semantic types of the head nouns modified by *hēi* “black”, etc. The discourse features are associated with the contextual information beyond phrasal structures, including the functional types of the clause containing *hēi* “black”, the mood, etc. A detailed description of all features with examples can be found in Appendix 1. Table 3 below provides a list of the features and their corresponding variable levels:

**Table 3.** List of the 46 feature types and corresponding feature levels

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Feature</th>
<th>Variable levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lexical-Collocational Information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>modifier</td>
<td>Negation Marker</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td>Degree Marker</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td><em>bèi</em> (被)</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td><em>yê</em> (也)</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td><em>qì</em> (起)</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td><em>le</em> (了)</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td><em>zhe</em> (着)</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td><em>dōu</em> (都)</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td><em>hài</em> (还)</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td><em>jiù</em> (就)</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td><em>guò</em> (过)</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td><em>râng</em> (让) / <em>líng</em> (令) / <em>shì</em> (使)</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td>Past Time Marker</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td>Future Time Marker</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td>Comparison Marker</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td>Frequency/Duration Marker</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td>Capability/Intention Marker</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td><em>yuè</em> (越) / <em>yù</em> (愈)</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>modifier</td>
<td>Quantifier</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>modifier</td>
<td>Doubt</td>
<td>2: yes/no</td>
</tr>
<tr>
<td><em>hēi</em>-<em>hēi</em></td>
<td>The frequency of <em>hēi</em> exceeds one time</td>
<td>3: no/yes diff M/yes same M</td>
</tr>
<tr>
<td>color terms</td>
<td>Cooccurrence with other color terms</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>noun phrase</td>
<td>Color object</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>notional word</td>
<td>parallel with <em>hēi</em></td>
<td>3: no/yes diff M/yes same M</td>
</tr>
<tr>
<td>compound word</td>
<td>phase marker</td>
<td>2: yes/no</td>
</tr>
</tbody>
</table>

### 2. Morphosyntactic & Semantic Information

<table>
<thead>
<tr>
<th>POS</th>
<th>黑 as noun</th>
<th>2: yes/no</th>
</tr>
</thead>
<tbody>
<tr>
<td>POS</td>
<td>黑 as verb</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>POS</td>
<td>黑 as adjective</td>
<td>3: no/yes-att./yes-pre.</td>
</tr>
<tr>
<td>lexicalization</td>
<td>lexicalized compound with 黑</td>
<td>4: yes-NP/yes-VP/yes-AdjP/no</td>
</tr>
<tr>
<td><em>hēi</em>-<em>hēi</em></td>
<td>Syntactic role of 黑</td>
<td>3: nominal/predicate/modifier</td>
</tr>
<tr>
<td>verb phrase</td>
<td>Syntactic role of VP containing 黑</td>
<td>6: null/sub./obj./pre./att./adv.</td>
</tr>
<tr>
<td>noun phrase</td>
<td>Syntactic role of NP containing 黑</td>
<td>5: null/sub./obj./att./adv.</td>
</tr>
<tr>
<td>noun phrase</td>
<td>Semantic type: dep-relation: sub</td>
<td>6: see notes</td>
</tr>
<tr>
<td>noun phrase</td>
<td>Semantic type: dep-relation: obj</td>
<td>6: see notes</td>
</tr>
<tr>
<td>noun phrase</td>
<td>Semantic type: dep-relation: iobj</td>
<td>6: see notes</td>
</tr>
<tr>
<td>noun phrase</td>
<td>Countability of NP containing 黑</td>
<td>3: mass/count/null</td>
</tr>
<tr>
<td>verb phrase</td>
<td>Semantic type: N collocated with 黑</td>
<td>6: see notes</td>
</tr>
<tr>
<td>noun phrase</td>
<td>Semantic type: 黑 as attribute</td>
<td>6: see notes</td>
</tr>
<tr>
<td>noun phrase</td>
<td>Semantic type: 黑 as head N</td>
<td>6: see notes</td>
</tr>
</tbody>
</table>

### 3. Discourse Information

<table>
<thead>
<tr>
<th>clause</th>
<th>Clause type</th>
<th>2: main/dependent</th>
</tr>
</thead>
<tbody>
<tr>
<td>clause</td>
<td>Types of dependent clause</td>
<td>3: relative clause/adv. clause/null</td>
</tr>
<tr>
<td>sentence</td>
<td>The omission of co-arguments with 黑</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>sentence</td>
<td>Pronouns</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>sentence</td>
<td>Explication of 黑</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>pragmatic</td>
<td>Whether 黑 is ambiguous</td>
<td>2: yes/no</td>
</tr>
<tr>
<td>pragmatic</td>
<td>Mood</td>
<td>3: see notes</td>
</tr>
</tbody>
</table>

Notes: the semantic types consist of abstract entity, body part, animate, inanimate object, organization, and null. The sentence mood consists of declarative, interrogative, and imperative.

### 3.3 Feature Annotation

The proposed 46 contextual features were manually annotated on 800 instances (100
for each sense), which produced a data frame of 105,600 data points. The annotation was done by two native speakers who are linguistically well-trained, and the inter-rater reliability was measured by calculating the unweighted Cohen’s Kappa coefficient. Given that the selected features are quite straightforward and detectable, a high degree of unweighted Cohen’s Kappa was obtained (0.9120, p = 0). The value is greater than 0.80, which indicates an almost perfect strength of agreement between the two annotators, according to Landis (1977), Altman (1990), and McHugh (2012). For items where the two coders' annotations differed, the two annotators discussed until they reached an agreement. After checking the consistency and reliability, the annotated data were used for the further BP analysis, as detailed in the next section.

4. Behavioral Profiles Analysis

After completing the manual annotation, a cluster analysis was conducted to identify the semantic (dis)similarities among the eight metaphorical meanings of hēi “black” based on the BP results. In the following, we first present a preliminary visualization of the distributional relations of the eight meanings based on the distance matrix. They were calculated from their contextual features (ID tags) with non-metric Multidimensional Scaling (MDS) algorithm. Then, the result of a hierarchical agglomerative cluster analysis (HAC) is presented to further explore the effectiveness of each contextual feature in differentiating the metaphorical meanings. In particular, the contextual features with a high effect on the grouping of metaphorical meanings are discussed in more detail to explore the underlying cognitive mechanisms of them.

4.1 General Observation on Cluster Solution

The initial findings are presented in Figure 1 below, which shows the preliminary
analysis of the (dis)similarities among the eight metaphorical meanings based on the contextual features. Specifically, the distance matrix was computed out of the annotated 800 instances based on ‘Gower distances’ (Gower, 1971). The distance matrix was then plotted with color labels for different meanings, based on the Kruskal non-metric MDS approach (Kruskal, 1964). It is found that even though the stress value is relatively high, the eight metaphorical meanings of *hēi* “black” can still be roughly divided into three clusters, located respectively at the left, the middle to the left, and the right of this plot. The left-most cluster consists of three metaphorical meanings: “Slander/Entrap-Blue”, “Network Attack-Green”, and “Unfavorable/Bad-Pink”, while “Unfavorable/Bad-Pink” has a broader distribution on the left side of the figure. The right-most cluster consists of another three metaphorical meanings: “Illegal/Underground-Dark Green”, “Secret/Mysterious-Dark Blue”, and “Unexpected/Surprising-Purple”. The left-of-center cluster mainly consists of two metaphorical meanings, i.e., “Angry/Sullen-Red” and “Evil/Malevolent-Brown”.

**Figure 1.** A two-dimensional Kruskal non-metric MDS of the eight metaphorical meanings of *hēi* “black” based on the 46 contextual features (stress = 28.87566, metric = Gower)
By visualizing the annotated instances in the two-dimensional plot, Figure 1 presents the semantic distribution of the eight proposed metaphorical meanings of *hēi* “black”. However, due to the limitation of MDS, we cannot interpret the location of the data points (the instances) directly with regard to the input contextual features. Hence, a HAC analysis was applied to further explore the underlying semantic relations among the different metaphorical meanings based on the annotated data. Specifically, the key BP vectors of the eight meanings, extracted from the annotated contextual information, were compared via ‘Canberra distance’ and the ‘ward.D2’ hierarchical clustering method.

The clustering results were visualized via a dendrogram, as shown in Figure 2. The generated clusters in the dendrogram are accompanied by three values, which provide validation for the corresponding cluster solution. It is noted that this validation technique was first used for language data in Divjak (2010). In Figure 2, the grey edge number shows the sequential order of the generation of each cluster. The remaining two values are the red-colored ‘Approximately Unbiased (AU) p-value’ and the green-colored ‘Bootstrap Probability (BP) value’, which measure the validations of each cluster solution. In general, the AU p-value is considered more reliable and precise for the validation of a cluster solution.
Figure 2. Cluster Dendrogram with AU/BP values (%) (distance = Canberra, method = ward.D2). Note that the closer the AU and BP value is to 1, the more empirical support the cluster has.

As indicated by the edge numbers, the first three generated clusters are (“Secret/Mysterious”, “Unexpected/Surprising”) ($p = 0.62$), (“Angry/Sullen”, “Evil/Malevolent”) ($p = 0.99$), and (“Network Attack”, “Slander/Entrap”) ($p = 0.90$). In the left branch at one-level up, “Illegal/Underground” is clustered with (“Unexpected/Surprising”, “Illegal/Underground”) ($p = 0.94$), and “Unfavorable/Bad” is clustered with (“Network Attack”, “Slander/Entrap”) ($p = 0.99$). At a higher level, (“Unfavorable/Bad”, “Slander/Entrap”, “Network Attack”) is clustered with (“Angry/Sullen”, “Evil/Malevolent”) ($p = 0.97$), forming a larger branch in the right side of the dendrogram. Finally, the two higher branches are linked to construct the whole dendrogram.

After obtaining the dendrogram, we examined the cluster solution of the eight metaphorical meanings from a hierarchical perspective. The concern is how to identify the optimal cluster solution, since there are several optional solutions based on different hierarchies. Hence, the average silhouette width was used to calculate the optimal number of clusters ($k$) and the corresponding cluster membership, which shows the
average well-formedness of the clusters in a given solution. With the help of *pvclust*, two similar silhouette width values were obtained when clustering the eight metaphorical meanings into two or three clusters, which are 0.2103 (k = 2) and 0.2072 (k = 3) respectively. For the two-cluster solution, one cluster contains the three meanings “Illegal/Underground”, “Unexpected/Surprising” and “Secret/Mysterious” (Cluster 1), and the remaining five meanings form another cluster (Cluster 2). For the three-cluster solution, Cluster 1 still contains “Illegal/Underground”, “Unexpected/Surprising” and “Secret/Mysterious”. The other two clusters are generated by grouping the two meanings “Angry/Sullen” and “Evil/Malevolent” as Cluster 2, and the remaining three meanings “Unfavorable/Bad”, “Network Attack” and “Slander/Entrap” as Cluster 3.

Given the subtle difference between the two silhouette width values (0.0031), we argue that exploring the three-cluster solution is reasonable, based on the MDS plot in Figure 1 and the observation that the Cluster-2 meanings demonstrate different usage patterns compared with the other three meanings in Cluster 3. In other words, it is necessary to explore the semantic (dis)similarities between Cluster 1 and Clusters 2 & 3, as well as between Cluster 2 and Cluster 3. As shown in Figure 2, the three clusters are marked with red frames based on the three-cluster membership, including Cluster 1: (“Unexpected/Surprising”, “Illegal/Underground”, “Secret/Mysterious”) (p = 0.94), Cluster 2: (“Angry/Sullen”, “Evil/Malevolent”) (p = 0.99), Cluster 3: (“Unfavorable/Bad”, “Network Attack”, “Slander/Entrap”) (p = 0.99).

In general, there are good reasons to believe that the same cluster solutions can be observed even if we use a different dataset, since the AU p-values of the three identified clusters are very strong (mean = 0.973, greater than 0.95). In the following, we focus on the between-cluster dissimilarities by comparing the effect size of the contextual features between Cluster 1 and Clusters 2 & 3, as well as those between Cluster 2 and Cluster 3, to examine which features play more critical roles in setting the clusters apart.
4.2 Behavioral Profiles: Cluster 1 vs. Clusters 2 & 3

This section explores the underlying semantic relations between Cluster 1 and Clusters 2 & 3 by examining the distinctive contextual features in the cluster solution, as was done in Divjak & Gries (2006). The distinctiveness of the contextual features in differentiating Cluster 1 and Clusters 2 & 3 was measured by two metrics: the Absolute Difference between the average BP vectors of the two groups, and Cramer’s V for each contextual feature in different groups. In the discussions, the effect size refers to the metric of Absolute Difference, and Cramer’s V is mainly used to verify it.

4.2.1 Overview of Distinctive Contextual Features
As shown in Figure 3 below, the snake plot (cf. Divjak & Gries, 2009) provides a list of contextual features ranked by their effect size. The higher the absolute value of the effect size is, the more significant the impact of the corresponding feature is on distinguishing Cluster 1 from Clusters 2 & 3. In other words, the features at the upper right and lower left are more distinctive in differentiating the two compared objects, while the features in the middle portion are less distinctive for the clustering.
Figure 3. Overview of the contextual features in Cluster 1 and Clusters 2 & 3, ranked by effect size.

To better demonstrate these contextual features with a high effect size, Table 4 shows an overview of the top 10 distinctive features with their effect size and Cramer’s V. Note that some features in Cluster 1 and Clusters 2 & 3 are overlapping due to the binary distinction of them.

Table 4. Top 10 features setting Cluster 1 and Clusters 2 & 3 apart based on the effect size

<table>
<thead>
<tr>
<th>Contextual feature</th>
<th>Variable level</th>
<th>Effect size</th>
<th>Cramer’s V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feature Type</td>
<td>Feature Value</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>---------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>SynR_CT</td>
<td>modifier</td>
<td>0.7653</td>
<td>0.757</td>
</tr>
<tr>
<td>POS_Adj</td>
<td>yes_attributive</td>
<td>0.7647</td>
<td>0.746</td>
</tr>
<tr>
<td>Countability_NP_CT</td>
<td>count</td>
<td>0.7553</td>
<td>0.782</td>
</tr>
<tr>
<td>SemT_sub</td>
<td>null</td>
<td>0.6167</td>
<td>0.606</td>
</tr>
<tr>
<td>POS_Verb</td>
<td>no</td>
<td>0.5567</td>
<td>0.565</td>
</tr>
<tr>
<td>Lexicalization</td>
<td>yes_NP</td>
<td>0.5173</td>
<td>0.563</td>
</tr>
<tr>
<td>SynR_NP_CT</td>
<td>object</td>
<td>0.444</td>
<td>0.747</td>
</tr>
<tr>
<td>SemT_dobj</td>
<td>null</td>
<td>0.336</td>
<td>0.399</td>
</tr>
<tr>
<td>SemT_NP_CT_attribute</td>
<td>abstract_entity</td>
<td>0.2867</td>
<td>0.773</td>
</tr>
<tr>
<td>SemT_NP_CT_attribute</td>
<td>animate</td>
<td>0.2713</td>
<td>0.773</td>
</tr>
</tbody>
</table>

### Clusters 2 & 3

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Feature Value</th>
<th>Value 1</th>
<th>Value 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SynR_NP_CT</td>
<td>null</td>
<td>-0.7733</td>
<td>0.747</td>
</tr>
<tr>
<td>Countability_NP_CT</td>
<td>null</td>
<td>-0.7733</td>
<td>0.782</td>
</tr>
<tr>
<td>SemT_NP_CT_attribute</td>
<td>null</td>
<td>-0.7647</td>
<td>0.773</td>
</tr>
<tr>
<td>SynR_NP_CT</td>
<td>predicate</td>
<td>-0.7067</td>
<td>0.757</td>
</tr>
<tr>
<td>POS_Adj</td>
<td>no</td>
<td>-0.5767</td>
<td>0.746</td>
</tr>
<tr>
<td>POS_Verb</td>
<td>yes</td>
<td>-0.5567</td>
<td>0.565</td>
</tr>
<tr>
<td>Lexicalization</td>
<td>no</td>
<td>-0.3688</td>
<td>0.563</td>
</tr>
<tr>
<td>SemT_sub</td>
<td>animate</td>
<td>-0.2853</td>
<td>0.606</td>
</tr>
<tr>
<td>le</td>
<td>yes</td>
<td>-0.2467</td>
<td>0.328</td>
</tr>
<tr>
<td>Quantifier</td>
<td>no</td>
<td>-0.2347</td>
<td>0.35</td>
</tr>
</tbody>
</table>

As indicated by features 1, 2, 5, 6 for Cluster 1 in the table, the three metaphorical meanings in Cluster 1 are always syntactically encoded as an adjectival modifier in a noun phrase, instead of being a verb. The frequent use of *hēi* “black” as an adjectival modifier in Cluster 1 leads to the absence of other syntactic roles collocating with the term as part of the predicate. This is demonstrated by features 4 and 8 in Cluster 1, namely, the absence of the semantic type of subject and object that collocate with *hēi* “black”. Examples (3)-(5) illustrate such uses of the attributive modifier.

(3) …仍然想拿黑[Illegal, Underground]收入，咋办?

*réngrán xiǎng ná hēi-shōurù, zà bàn?*

still want take black income how do
(“What should (we) do if (they) still want to take the illegal income?”)

(4) …搭载多项黑[Secret, Mysterious]科技。

dāzài duō-xiàng hēi-kējì

carry many-CL black technology

(“Equipped with a number of black technologies.”)

(5) …吉林队是历届联赛中最黑的一匹黑[Unexpected, Surprising]马。

Jílín duì shì lì-jìè liánsài zhōng zuì hēi-de yī-pǐ hēi-mǎ

Jilin team BE all-previous league in most black-DE1 one-CL black horse

(“The Jilin team is the most unexpected black-horse in previous leagues.”)

In addition, features 3, 9, and 10 for Cluster 1 indicate that the noun phrases containing hēi are more likely to refer to countable notions or physical entities. This can be illustrated by the notional entities hēi-shōurù “black income” and hēi-kējì “black technology”, and the physical entity hēi-mǎ “black horse” in the above examples. Due to their countability, such noun phrases may collocate with a quantifier, such as duō-xiàng “several-Classifier” with “technology”, and yī-pǐ “one-Classifier with hēi-mǎ “black horse”. Related to that, feature 7 for Cluster 1 further describes the syntactic role of these noun phrases, as they are frequently used as the object. In total, there are five contextual features directly related to the noun phrases containing hēi “black”, which include the part of speech of hēi “black”, the syntactic roles of hēi “black” in the NP, the countability of the NP, the semantic types of the NP, and the syntactic roles of the NP. Overall, the distinctive features for Cluster 1 reveal that the NPs containing hēi “black” play a vital role in distinguishing Cluster 1 (“Illegal/Underground”, “Secret/Mysterious” and “Unexpected/Surprising”) from other two clusters.

For Clusters 2 & 3, many of the distinctive contextual features overlap with those appearing in Cluster 1, including features 1-8. In view of the fact that these features pertain to the attributive vs. predicative usage patterns of hēi “black”, it is not
surprising to find that the set of features are effective in both groups, but work with opposite effects in different cluster groups, given the way in which the effect size was calculated. That is to say, the two groups may be dichotomous in relation to these features. For example, opposite to the tendency of adjectival uses in Cluster 1, features 1, 2, 3, and 5 indicate that it is comparatively rare to find the instances of hēi “black” in Clusters 2 & 3 occurring as an adjectival or attributive modifier. Moreover, features 4 and 6 for Clusters 2 & 3 show that hēi “black” is more frequently used as a verbal predicate when referring to the meanings in the group. In terms of semantic types, feature 8 shows that the subjects of hēi “black” tend to be “Animate”, which refer to proper names or pronouns in most cases. In terms of aspectual features, feature 9 indicates that the predicative uses of hēi “black” tend to co-occur with the aspectual marker le, denoting the realization of an event or state. The predicative uses of hēi “black” are illustrated below.

(6) …她无意中黑[Slander, Entrap]了[Aspect Marker]青岛一把。
    tā wúyìzhōng hēi le Qīngdāo yī-bā
    she unintentionally black LE Qingdao one-CL
    ("She unintentionally blackened Qingdao one time.")

(7) 穆司爵脸黑[Angry, Sullen]了[Aspect Marker]，其他人都忍不住笑了。
    Mùsījué liǎn hēi le, qítā rén dōu rěn-bù-zhù xiào le
    NAME face black LE, other person all cannot-help laugh LE
    ("Mu Sijue’s face blackened, while the others couldn’t help laughing.")

(8) 央视官网被黑[Network Attack]达两小时。
    Yang-shí guān-wǎng bèi hēi dá liǎng xiāoshí
    Central-Television official-website BEI black to two hours
    ("CCTV’s official website was hacked for two hours.")
Given the predominant predicative uses of hēi “black” in Clusters 2 & 3, it is not surprising to see that two of the effective features in the group are the absence of a lexicalized NP and the lack of quantifier, as shown by features 7 and 10 for Clusters 2 & 3 in Table 4. The top 10 features indicate that Cluster 1 can indeed be separated from Clusters 2 & 3, as they display quite distinct usage patterns at the collocational and morphosyntactic levels. Based on that, the next section discusses the cognitive mechanisms that help motivate the different groupings of meanings.

4.2.2 Cognitive Semantic Analysis Based on the Distinctive Contextual Features

Color is, in general, a salient property of physical objects. Color perception is one type of salient human experience via visionary reception for the wavelengths of the light reflected from a physical object. The perceptual associations between colors and non-visual entities are well demonstrated by the BP results of this study. In Cluster 1, the visual perception of hēi “black” is used as a modifier for non-visual evaluation associated with the color. As shown by the distinctive features, the three metaphorical meanings in Cluster 1 (“Illegal/Underground”, “Secret/Mysterious”, “Unexpected/Surprising”) mainly occur in a noun phrase, via compounding hēi “black” with a lexicalized noun. In such uses, hēi “black” is an attributive of a nominal entity. The conceptual metaphor underlying Cluster 1 can be generally summarized as “NON-VISUAL EVALUATION (or VALUE JUDGEMENT) IS COLOR PERCEPTION”, where the black color provides a conceptual source for describing the non-color notions. The perceptual experience “the color of an object is black” from the source domain offers the conceptual base for the extended meanings in Cluster 1. In other words, the three metaphorical meanings of hēi “black” in Cluster 1 are extended from “blackness in color” to “darkness in quality or value judgement”. Based on this conceptual extension, the color sense “blackness” is mapped onto the invisible, evaluative target domain to describe illegal/unacceptable deeds, secretive/unknown information, or surprising/unexpected quality, as they are cognitively related and grouped in Cluster 1.

In conclusion, the conceptual metaphor witnessed in Cluster 1 is: “AN
UNPLEASANT OR UNEXPECTED QUALITY IS BLACK”, which entails the following correspondences in its conceptual mappings: “The quality of a person (entity) corresponds to the color of a physical object”, “The valuation (judgement) on the quality corresponds to the perception of the color”, and “The unpleasantness or unexpectedness of that quality corresponds to the black color”. It follows the general principle of Conceptual Metaphor Theory whereby a more familiar, visible, concrete, or bodily experience (vision) is used to describe less familiar, invisible or non-concrete experiences (quality or values).

Figure 4. The conceptual base for the metaphorical meanings in Cluster 1

In contrast to the attributive uses in Cluster 1, hēi “black” in Clusters 2 & 3 with the other five metaphorical meanings tends to be used syntactically as a verbal predicate, and the semantic types of its subjects are more likely to be animate. The verbal uses of hēi “black” profile a process or activity of color change (turning black) rather than attributing the static color of an entity. Thus, there involves a different conceptual metaphor in Clusters 2 & 3, which can be specified as “A CHANGE OF NON-VISUAL STATE, RELATION, OR STATUS IS A CHANGE OF COLOR”, and more specifically, “TURNING INTO AN UNPLEASANT STATE OR RELATION IS TURNING BLACK”. Figure 5 below represents the common conceptual base for the five meanings in Clusters 2 & 3; that is, “the color of a physical object turns black”. The conceptual base refers to the visual experience of seeing something “turning black”, which is extended to the non-visual target domain to describe an unpleasant change of internal state or interpersonal relation. In other words, the entailed conceptual mappings can be summarized as: “The internal state or interpersonal relation corresponds to the color of a physical object”, and “The unpleasant change of the internal state or interpersonal
relation corresponds to the process of turning black in color of the physical object”.

![Diagram](image)

**Figure 5.** The conceptual base for the metaphorical meanings in Clusters 2 & 3

In terms of cognitive motivations, the perception of color (Cluster 1) and the process of color change (Clusters 2 & 3) provide the two source domains for metaphorical extensions of *hēi* “black”, resulting in its polysemous meanings. When referring to the three attributive meanings in Cluster 1, *hēi* “black” describes the undesirable or unexpected properties of some objects; but when referring to the predicative meanings in Clusters 2 & 3, it describes the unpleasant changes of state or status. Having laid down the distinct conceptual metaphors involved in Cluster 1 and Clusters 2 & 3, the next section focuses on the differences between Cluster 2 and Cluster 3 to provide a more refined semantic analysis based on the BP results.

4.3 Behavioral Profiles: Cluster 2 vs. Cluster 3

This section further explores the semantic (dis)similarities between the meanings in Cluster 2 (“Angry/Sullen” and “Evil/Malevolent”) and those in Cluster 3 (“Unfavorable/Bad”, “Network Attack”, and “Slander/Entrap”). The distinctiveness of the annotated contextual features was measured by the Absolute Difference and the Cramer’s V between the two Clusters as in the previous section. Then, a cognitive semantic analysis was conducted based on the top 10 distinctive features for separating Cluster 2 and Cluster 3.

4.3.1 *Overview of Distinctive Contextual Features*
It is found that Cluster 2 and Cluster 3 share most of the distinctive contextual features, but the rankings and effects of the shared features are varied. Syntactically, Cluster 2 pertains to the uses of *hēi* “black” as an intransitive or stative predicate, while Cluster 3 pertains to its uses as a transitive predicate. Figure 6 below shows the distribution of all contextual features with a snake plot view, and Table 5 lists the top 10 distinctive features for Cluster 2 and Cluster 3 respectively.

![Figure 6](image.png)

**Figure 6.** Overview of contextual features in Cluster 2 and Cluster 3, ranked by effect size.

It is shown from Figure 6 that the most distinctive features are distributed at both ends of the snake plot. The average absolute value of the effect size of the top 10 contextual features is 0.2968, which is lower than the value between Cluster 1 and Clusters 2 & 3.
(0.5298). This indicates that the discrimination between Cluster 2 and Cluster 3 is lower or less significant in the dataset. It is important to note that even though the top 10 contextual features are similar in Cluster 2 and Cluster 3, the differences in their rankings of the effect size and variable levels can still provide informative evidence for distinguishing the two clusters. Table 5 below lists the top 10 distinctive features in the two clusters:

Table 5. The top 10 features setting Cluster 2 and Cluster 3 apart based on effect size

<table>
<thead>
<tr>
<th>Contextual feature</th>
<th>Variable level</th>
<th>Effect size</th>
<th>Cramer’s V</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cluster 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-1. SemT_dobj</td>
<td>null</td>
<td>-0.535</td>
<td>0.568</td>
</tr>
<tr>
<td>2-2. SemT_sub</td>
<td>body_part</td>
<td>-0.495</td>
<td>0.626</td>
</tr>
<tr>
<td>2-3. POS_Adj</td>
<td>yes_predicative</td>
<td>-0.47</td>
<td>0.56</td>
</tr>
<tr>
<td>2-4. Omission_Co-Arguments</td>
<td>no</td>
<td>-0.40333</td>
<td>0.449</td>
</tr>
<tr>
<td>2-5. bei</td>
<td>no</td>
<td>-0.33</td>
<td>0.406</td>
</tr>
<tr>
<td>2-6. Deg_Marker</td>
<td>yes</td>
<td>-0.245</td>
<td>0.359</td>
</tr>
<tr>
<td>2-7. POS_Verb</td>
<td>no</td>
<td>-0.225</td>
<td>0.222</td>
</tr>
<tr>
<td>2-8. Lexicalization</td>
<td>no</td>
<td>-0.2221</td>
<td>0.387</td>
</tr>
<tr>
<td>2-9. SynR_CT</td>
<td>predicate</td>
<td>-0.2083</td>
<td>0.271</td>
</tr>
<tr>
<td>2-10. SynR_VP_CT</td>
<td>null</td>
<td>-0.1933</td>
<td>0.381</td>
</tr>
<tr>
<td><strong>Cluster 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-1. Omission_Co-Arguments</td>
<td>yes</td>
<td>0.40333</td>
<td>0.449</td>
</tr>
<tr>
<td>3-2. POS_Adj</td>
<td>no</td>
<td>0.3417</td>
<td>0.56</td>
</tr>
<tr>
<td>3-3. bei</td>
<td>yes</td>
<td>0.33</td>
<td>0.406</td>
</tr>
<tr>
<td>3-4. SemT_sub</td>
<td>null</td>
<td>0.3083</td>
<td>0.626</td>
</tr>
<tr>
<td>3-5. Deg_Marker</td>
<td>no</td>
<td>0.245</td>
<td>0.359</td>
</tr>
<tr>
<td>3-6. POS_Verb</td>
<td>yes</td>
<td>0.225</td>
<td>0.222</td>
</tr>
<tr>
<td>3-7. SemT_dobj</td>
<td>abstract_entity</td>
<td>0.21333</td>
<td>0.568</td>
</tr>
<tr>
<td>3-8. SemT_N_Collocated_CT</td>
<td>null</td>
<td>0.1933</td>
<td>0.393</td>
</tr>
<tr>
<td>3-9. POS_Noun</td>
<td>yes</td>
<td>0.18</td>
<td>0.273</td>
</tr>
<tr>
<td>3-10. Lexicalization</td>
<td>yes_NP</td>
<td>0.1767</td>
<td>0.387</td>
</tr>
</tbody>
</table>

Table 5 shows that the part of speech (POS_Adj) of "hēi “black” (i.e., whether it is an adjective or not) is one of the most distinctive features for both Cluster 2 and Cluster 3. Specifically, features 2-3, 2-7, and 2-9 indicate that "hēi “black” in Cluster 2
prototypically occurs as a stative predicate with a degree adverb (example 9), not necessarily as an intransitive verb (example 7), in the dataset. In terms of the argument structure, features 2-1 and 2-5 manifest the absence of the direct objects and the absence of the passive marker bèi, clearly indicating low transitivity in Cluster 2. In terms of the collocating subjects, feature 2-2 shows that the semantic types of the subjects tend to be “Body Part” in Cluster 2. Example 9 is a good illustration of this usage:

(9) 叶咎的脸很黑。

Yèjiù-de liǎn hěn hēi
NAME-DE₁ face quite black
(“Ye Jiu's face is quite black.”)

In addition, features 2-6 and 2-8 provide more information about the morphosyntactic behaviors of hēi “black” as a stative predicate, not a pre-nominal modifier. Feature 2-6 indicates that hēi “black” in Cluster 2 is more likely to co-occur with a degree marker, such as the default degree adverb hěn “pretty, very”, or other comparative markers, such as tái “too” or bìjiào “comparatively”. The frequent collocation with degree markers clearly manifests scalarity or gradability in the uses of hēi “black” in Cluster 2. Feature 2-8 indicates that hēi “black” is less likely to be compounded within a noun phrase, but more likely to be associated with intransitivity. Moreover, feature 2-10 shows that hēi “black” in Cluster 2 is less likely to appear in a VP. Example 10 below serves to illustrate such tendencies:

(10) 你的心太黑了，连十来岁的孩子都不放过。

nǐde xīn tài hēi le, lián shí-lái-suí-de hǎi-zi dōu bù fāngguò
your heart too black LE, even teenage-DE₁ child all BU let-go
(“Your heart is so dark that you can’t let go of even teenagers.”)
In comparison, Cluster 3 shows a tendency of verbal uses. When referring to the meanings in Cluster 3 (“Unfavorable/Bad”, “Network Attack”, “Slander/Entrap”), hēi “black” is more likely to be a verbal, transitive predicate, rather than stative predicate, according to features 3-2, 3-6 and 3-7. Related to the verbal use, it is not surprising to find that the collocating nouns with hēi “black” are mostly either the subjects or the direct objects in the sentences, as revealed by feature 3-8. Below is a typical example of the transitive use, which is characteristic of Cluster 3:

(11) 郭文贵在黑[Slander, Entrap]了刘志华后保住盘古大观…

Guō-Wénguì zài hēi le Líú-Zhihuá hòu bāozhù Pángǔ-dàguān
NAME at blacken LE NAME after retain Pangu-Plaza
(“Guo Wengui kept the Pangu Plaza after blackening Liu Zhihua…”)

As hēi “black” is used predominantly as a transitive-causative verb in Cluster 3, it frequently collocates with the passive marker bèi, but rarely with a degree marker, as indicated by features 3-3 and 3-5 and exemplified below:

(12) 黄子韬从出道以来一路被[Passive Marker][Assailant omitted]黑[Slander, Entrap]。

Huáng-zǐtāo cóng chūdào yǐlái yīlù bèi hēi
NAME from debut since all-the-way BEI slander
(“Huang Zitao has been slandered all the time since his debut.”)

(13) 萧敬腾经纪公司网页被[Passive Marker][Assailant omitted]黑[Network Attack]。

Xiāo-jìngténg jīngjì-gōngsī wǎngyè bèi hēi
NAME talent-agency webpage BEI hack
(“The website of Xiao Jingteng’s talent agency was hacked.”)
On the other hand, features 3-9 and 3-10 indicate that *hēi* “black” in Cluster 3 may occur as a noun or lexicalize in an NP with other words, which shows that the morphosyntactic behaviors of *hēi* “black” in Cluster 3 are more flexible than what are observed in Cluster 1 and Cluster 2, especially when it refers to the sense “Unfavorable/Bad”. This also helps to explain the broader distribution of Cluster-3 instances in the MDS plot (Figure 1). In sum, *hēi* “black” in Cluster 3 tends to be a verb, but it can also be nominalized (as in example 14), which is used as a modifier (as in example 15), or as the R-component in a V-R (Verb-Result) compound (as in example 16).

(14) …给自己的人生重重地抹了一笔黑[Unfavorable, Bad: Noun]。

    gěi zìjǐde rénshēng chóngchóngde mǒ-LE yī-bī hēi

    give oneself ’s life heavily smear-LE one-CL black

    (“…heavily smeared their own lives.”)

(15) 你家爱豆的童年黑[Unfavorable, Bad: Adjective Modifier]历史已上线。

    nǐ-jīā àidòu-de tóngnián hēi-lǐshǐ yǐ shàng-xiàn.

    Your-family idol-DE childhood black history already on-line

    (“The black history of your idol’s childhood has been publicized online.”)

(16) 你想被[Assailant omitted]拉黑[Unfavorable, Bad: (V)-R]吗?

    nǐ xiǎng bèi lā-hēi ma

    you want BEI pull-black MA

    (“Do you want to be blacklisted?”)

Moreover, regarding the omission of co-arguments, as indicated by the opposing features 2-4 and 3-1, the dissimilarities between Clusters 2 and 3 can be further verified at the discourse level. It is found that the collocated arguments of *hēi* “black” in the instances of Cluster 3 are more likely to be omitted, especially for the agentive subjects,
as indicated by feature 3-4. This is different from the instances occurring in Cluster 2. The omission of the co-arguments frequently found in Cluster 3 can be demonstrated by examples (12), (13) and (16) above, as the agentive arguments referring to [Assailant] in the event are omitted in all three examples.

4.3.2 Cognitive Semantic Analysis Based on Distinctive Contextual Features

As mentioned above, the five metaphorical meanings of ʰɛi “black” in Clusters 2 & 3 share the same conceptual base, i.e., “change of color”. Based on the source experience, different meanings of ʰɛi “black” are realized through metaphorical extensions accompanied by various morphosyntactic forms, collocational patterns, and other contextual features. This section further analyzes the semantic (dis)similarities between the two meaning groups from a cognitive semantic perspective.

![Diagram](image.png)

**Figure. 7** The conceptual base for the metaphorical meanings in Cluster 2

Considering the distinctive features for Cluster 2, the uses of ʰɛi “black” referring to “Angry/Sullen” or “Evil/Malevolent” describe a process of internal change, which is based on the source experience of a perceptual change in color. That is, the perceptual experience, “the blackening process in color change” (physical change) from the source domain, is extended to describe a non-visual or abstract process, “the darkening of emotional state or internal quality” (non-physical change). The shelled circles of the color change in Figure 7 schematically represent the process of the internal change (inside the shell). The target domains of the two meanings in Cluster 2 are respectively “turning angry/sullen in the emotional state” and “turning evil/malevolent in the mental state”, which are conceived as similar to the process of color change “turning black in color”. In other words, the conceptual metaphor realized in Cluster 2 can be
summarized as “AN UNPLEASANT CHANGE OF THE INTERNAL STATE IS AN OBJECT OR PERSON TURNING BLACK”, of which the entailed conceptual mappings are: “The emotional (mental) state corresponds to the color of a physical object”, and “Turning angry (sullen) emotionally or turning evil (malevolent) mentally corresponds to turning black in color of the physical object”. Based on that, hēi “black” in Cluster 2 is metaphorically used to evaluate or depict the unpleasant changes of the internal state of a person or entity.

Figure. 8 The conceptual base for the three meanings in Cluster 3. (Note that the Causer in the dotted-line frame means it may be omissible).

In the same vein, the perceptual experience for the three meanings in Cluster 3 (“Unfavorable/Bad”, “Network Attack”, “Slander/Entrap”) can be described as “a causative act initiated by a Causer to turn a person or entity black”. Figure 8 above schematically represents a causative process of color change in between the two layers of circles, to profile the changes of external or social behaviors, not the internal quality. The experience of the causative change is extended to describe “any assailing act that blackens the status or reputation of a person or organization”. Therefore, the conceptual metaphor for the three meanings in Cluster 3 can be specified as: “AN ACT THAT (INTENTIONALLY) HARMS OR EXCLUDES A PERSON OR ENTITY IS AN ACT OF MAKING A PERSON OR ENTITY BLACK”. The entailed conceptual mappings in that metaphor include: “The status or reputation of a person (organization) corresponds to the color of a physical object”, and “Harming someone’s status or reputation corresponds to blackening the physical object”. Based on such a conceptual metaphor, hēi “black” in Cluster 3 is metaphorically used to describe the deteriorating
changes or exclusions in social interaction or interpersonal relationship between entities by an external Causer. The initiators of such assailing act may be unknown or omitted ([Assailant] omitted), as seen in features 3-1 and 3-4, but the patients (or “Victim”) of the blackening act are often present, since they are the most salient participants in the event, from the perspective of ‘Frame Semantics’ (Fillmore, 1985). It is also noted that the meaning of “Unfavorable/Bad” in Cluster 3 is further derived to describe the resultant state of the act. The use of hēi “black” as a resultative complement may be less prototypical in its verbal usage since it can only occur in a V-R compound to be combined with another active verb, as seen in lā-hēi “pull-black, blacklist”, jiā-hēi “add-black, blacklist”, mǒ-hēi “smear-black, blacken”.

In sum, it is pretty rewarding to see that the BP results obtained in this study can be accounted for with further consideration of the cognitive semantic motivations. The correlations of the morphosyntactic forms and the lexical-semantic distinctions in the metaphorical uses of hēi “black” are quite striking, as discussed above. The grouping of polysemous senses is ultimately triggered by the conceptual metaphors that underlie the varied metaphorical extensions.

5. Conclusion

In this paper, we have explored the semantic relations between the eight metaphorical meanings of hēi “black” with the BP approach to identify the distinctive features associated with the varied senses. With a dataset consisting of 800 instances (100 for each sense) annotated with 46 contextual features, it is found that the eight metaphorical meanings of hēi “black” can be categorized into three clusters based on the proposed contextual features (ID tags) in relation to their lexical-grammatical behaviors, other semantic features, as well as discourse information. According to the hierarchical agglomerative cluster analysis, the three clusters are: Cluster 1 with three related meanings (“Unexpected/Surprising”, “Secret/Mysterious”, “Illegal/Underground”),
Cluster 2 with two related meanings (“Angry/Sullen”, “Evil/Malevolent”), and Cluster 3 with three related meanings (“Unfavorable/Bad”, “Network Attack”, “Slander/Entrap”). The three meanings in Cluster 1 manifest the distinctive behaviors of hēi “black” occurring with contextual features pertaining to a noun phrase, as the instances of hēi “black” in Cluster 1 tend to be a pre-nominal modifier attributing the following noun. The behaviors of hēi “black” in Cluster 1 are comparatively more consistent with fewer contextual variations than the other five meanings. hēi “black” in Cluster 2 behaves as a stative predicate in an intransitive construction, describing a gradable change of state pertaining to the subjects, which are often instantiated by a “BODY PART”. For Cluster 3, hēi “black” is prototypically used as a transitive verb in the transitive-causative construction, with more flexible contextual variations.

Furthermore, the results of the Behavior Profiles are accountable from the perspective of Conceptual Metaphor Theory. The underlying cognitive semantic mechanisms for the different groupings of metaphorical senses were discussed with reference to their conceptual metaphors. It is proposed that the common experiences of “color perception” and related “change of color” serve as two source domains to conceptualize two non-perceptual target domains through metaphorical extensions. Respectively, the black color is used for “non-visual evaluation (or value judgement)”, and the color change of turning black is used for “change of non-visual state or status”. Precisely, the conceptual metaphors at work in the three clusters can be presented as: “AN UNPLEASANT OR UNEXPECTED QUALITY IS BLACK” (for Cluster 1); “AN UNPLEASANT CHANGE OF THE INTERNAL STATE IS AN OBJECT OR PERSON TURNING BLACK (for Cluster 2); “AN ACT THAT (INTENTIONALLY) HARMs OR EXCLUDES A PERSON OR ENTITY IS AN ACT OF MAKING A PERSON OR ENTITY BLACK” (for Cluster 3).

It is concluded that the eight metaphorical meanings of hēi “black” are well distinguished by means of comparing and profiling their usage patterns, including the collocational, morphosyntactic, semantic, and discourse features, which are correlated with the varied semantic extensions. The study ultimately demonstrates the
effectiveness of the corpus-based BP approach in the investigation of metaphorical polysemy. The proposed cognitive mechanisms for metaphorical extensions are well-motivated and realized in line with the Conceptual Metaphor Theory.

Endnotes

1. The capitalized gloss represents the function words in Chinese. Appendix 2 provides a list of the function words used in the cited examples.


3. The white color, 白 “white”, is another earliest-acquired CT in Chinese. It is noted that some metaphorical meanings of that term are not symmetrical with 黑 “black”. For example, there is no correspondences for the contrast in expectedness between the two CTs. That is, 黑[Unexpected] horse does not find an antonym, 白[Expected] horse, in Chinese, which calls for another research on the metaphorical polysemy of the Chinese CT 白 “white”.

4. In principle, the higher the stress value is, the more missing information there may be.

References


6. Appendix 1

6.1 Collocational Information
This section illustrates the contextual features focusing on collocational information.

6.1.1 Modifier: Negation Marker

Whether 黑 “black” collocated with a negation marker: 不, 没.

e.g., 不黑 不hei (“not black”)
6.1.2 Modifier: Degree Marker
Whether  “black” collocated with a degree marker: hěn 很; wánquán 完全; quánmiàn 全面; zhēn 真; tài 太; zhème 这么; nàme 那么; tèbié 特别; chèdǐ 彻底, et al.
   e.g., 太黑 tài hēi (“too black”) 

6.1.3 Modifier: bèi
Whether “black” collocated with passive marker bèi 被.
   e.g., 官网被黑 guānwǎng bèi hēi (“the website is hacked.”)

6.1.4 Modifier: yě
Whether hēi “black” collocated with yě 也 “also; as well; too”.
   e.g., 也黑 yě hēi (“also be black”)

6.1.5 Modifier: qǐ
Whether hēi “black” collocated with qǐ 起 “up”.
   e.g., 黑起来 hēi qǐlái (“turn black”)

6.1.6 Modifier: le
Whether hēi “black” collocated with le 了 “finish; realize; complete”.
   e.g., 被黑了 bèi hēi le (“be blackened”)

6.1.7 Modifier: zhe
Whether hēi “black” collocated with aspect particle zhe 着.
   e.g., 黑着脸 hēi zhe liǎn (“black face”)

6.1.8 Modifier: dōu
Whether hēi “black” collocated with dōu 都 “all; even; just”.
e.g., 比高利贷都黑  bǐ gāolìdài dōu hēi (“black than usury”)

6.1.9 Modifier: hái
Whether hēi “black” collocated with hái 还 “still; also”.
   e.g., 比墨还黑  bǐ mò hái hēi (“blacker than ink”)

6.1.10 Modifier: jiù
Whether hēi “black” collocated with jiù 就 “at once; already”.
   e.g., 就被黑了  jiù bèi hēi le (“got hacked”)

6.1.11 Modifier: guò
Whether hēi “black” collocated with guò 过 “once”.
   e.g., 被黑过  bèi hēi guò (“have been hacked”)

6.1.12 Modifier: ràng/líng/shǐ
Whether hēi “black” collocated with causative marker ràng 让/ líng 令/ shǐ 使.
   e.g., 让脸黑得像个煤球  ràng liǎn hēi-dè xiàng gè méiqióu (“make the face as black as a briquette”)

6.1.13 Modifier: Past Time Marker
Whether hēi “black” collocated with a past time modifier: cèng 曾, yǐ 已, yǐjīng 已经.
   e.g., 已经被黑  yǐjīng bèi hēi (“already be hacked”)

6.1.14 Modifier: Future Time Marker
Whether hēi “black” collocated with a future time modifier: jiāng 将, jiānghuì 将会, hui 会.
   e.g., 后续会黑化  hòuxù hui hēi-huà (“will turn black in the future”)

6.1.15 Modifier: Contrast
Whether hēi “black” collocated with comparison marker: bǐ 比, bǐjiào 比较, zuì 最, gèng 更.
   e.g., 更黑 gèng hēi (“blacker”)

6.1.16 Modifier: Frequency/Duration
Whether hēi “black” collocated with Frequency/Duration Marker: jīngcháng 经常, cì 次, yīrán 依然, shùnjiān 瞬间.
   e.g., 瞬间一黑 shùnjiān yī hēi (“turn black instantly”)

6.1.17 Modifier: Capability/Intention
Whether hēi “black” collocated with a modifier of capability or intention marker: yào 要, néng 能, xiǎng 想.
   e.g., 不不要太黑 bù yào tài hēi (“not too black”)

6.1.18 Modifier: yuè/yù
Whether hēi “black” collocated with yuè 越/yù 愈 “increasingly”.
   e.g., 越来越黑 yuè lái yuè hēi (“getting black”)

6.1.19 Modifier: Quantifier
Whether the noun phrase containing hēi “black” collocated with a quantifier: yī-xiàng 一项, yī-chǎng 一场, yī-pī 一批, et al.
   e.g., 一项黑科技 yī-xiàng hēi-kējì (“a black technology”)

6.1.20 Modifier: Doubt
Whether hēi “black” collocated with a marker of doubt: shìfǒu 是否, quèdìng 确定, nán guài 难怪, guòrán 果然, et al.
   e.g., 是否会一黑到底 shìfǒu huì yī hēi dào dǐ (“Whether it will be black to the end”)
6.1.21  

6.1.21 hēi “black”: The frequency of hēi “black” exceeds one time

Whether the frequency of occurrence of hēi “black” exceeds one time in an instance, if so, whether they refer to same meaning.

i. Exceeds one time but refers to same meaning:
   e.g., 李庭芳结识了这两名黑[Illegal, Underground]社会分子，听说了一些黑[Illegal, Underground]帮大佬的威风史。

   Li-tīngfāng jiéshì le zhè liǎng-míng hēi-shèhuì fēnzǐ, tīngshuō le yīxiē hēibāng
dàlǎo de wēifēng-shǐ

   ("Li Tingfang got acquainted with these two gangsters and heard about the prestigious history of several gang bosses.")

ii. Exceeds one time but refers to different meaning:
   e.g., 这匹黑[Unexpected, Surprising]马肤色并不黑[Black Color]，但他
   的成绩可够黑[Unexpected, Surprising]的。

   zhè-pǐ hēi-mǎ fūsè bù hēi, dàn tā de chéngjī kě gòu hēi de

   The skin color of this black horse is not black, but his grades are black enough.

6.1.22 Color Terms: Co-occurrence with other color terms

Whether hēi “black” co-occurred with other color terms in an instance.

   e.g., 这样的干部黑了心，红[red]了眼。

   zhèyàng-de gànbù hēi le xīn, hóng le yǎn

   ("The hearts of such cadres have turned black, and their eyes have turned red.")

6.1.23 Noun phrase: Color object

Whether hēi “black” co-occurred with a color object referring to the black color.

   e.g., 黑得像茄子 hēi-dé xiàng qiézi “as black as eggplant”
6.1.24 Notional word: Parallel with hēi “black”

Whether hēi “black” collocated with a notional word that parallel with it, if so, whether they refer to similar meaning.

i. Notional word referring to similar meanings with hēi “black”:

  e.g., 涉黑涉恶 shè hēi shè è (“involved black and wickedness”)

ii. Notional word referring to different meanings with hēi “black”:

  e.g., 不黑不吹 bù hēi bù chuī (“not slander either boast”)

6.1.25 Compound word: phase marker

Whether hēi “black” collocated with a phase marker in a compound word: chéng 成 “become”, diào 掉 “drop”, dào 到 “arrive”, jìn 进 “enter”, et al.

  e.g., 黑进他们的飞船 hēi-jìn tāmende fēichuán (“hack-into their spaceship”)

6.2 Syntactic Information

This section illustrates the contextual features focusing on syntactic information.

6.2.1 POS: Noun

Whether the part of speech of hēi “black” is noun.

  e.g., 抹了一笔黑 mó yī-bǐ hēi (“smeared a black stroke”)

6.2.2 POS: Verb

Whether the part of speech of hēi “black” is verb.

  e.g., 黑别人 hēi biérén (“blacken others”)

6.2.3 POS: Adjective

Whether the part of speech of hēi “black” is adjective, if so, whether it is attributive or
predicative.

i. Attributive adjective: 黑一日游 hēi yīrì-yóu (“black one-day trip”)

ii. Predicative adjective: 心很黑 xīn hěn hēi (“heart is very black”)

6.2.4 Lexicalization: hēi “black” composites with other lexical items
Whether hēi “black” lexicalized compound with other lexical items as a NP, VP or AdjP.

i. NP: 黑科技 hēi-kējì (“black technology”)

ii. VP: 拉黑 lā-hēi (“blacklist”)

iii. AdjP: 腹黑 fù-hēi (“scheming”)

6.2.5 hēi “black”: Syntactic role
Whether the syntactic role of hēi “black” is of type:

i. Nominal:
   e.g., 招来有心人的黑 zhāolái yǒuxīnrén de hēi (“incur the slander of conscientious”)

ii. Predicate:
   e.g., 朱潘再一次黑了薛老的邮箱 Zhūpān zàiyīcì hēi le Xuēlǎo de yóuxiāng
   (“Zhu Pan hacked Xue Lao’s mailbox again.”)

iii. Modifier:
   e.g., 黑油井的存在 hēi-yóujǐng de cúnzài (“the existence of black oil wells”)

47
6.2.6 Verb Phrase: Syntactic role

Whether the syntactic role of VP containing  
“black” is of type:

i. Subject:

   e.g., 自黑字面上是自己黑自己
   zi-hēi zìmiànshàng shì zìjǐ hēi zìjǐ
   (“Self-deprecation literally means someone deprecate themselves.”)

ii. Object:

   e.g., 避免出现自黑
   bìmǐăn chūxiàn zìhēi (“avoid self-deprecation”)

iii. Predicate:

   e.g., 拉黑了他的微信
   lā-hēi le tāde wéixìn (“blacklisted his WeChat”)

iv. Attributive:

   e.g., 自黑文
   zì-hēi wén (“self-deprecating pots”)

v. Adverb:

   e.g., 这两人黑着脸吵了起来
   zhè liǎn-rén hēi zheliăn chǎo le qǐ lái
   (“The two begin to quarrel with black faces.”)

6.2.7 Noun phrase: Syntactic role

Whether the syntactic role of NP containing  
“black” is of type:

i. Subject:

   e.g., 黑历史已上线
   hēi-lìshī yǐ shàngxiàn (“the black history has been online”)

48
ii. Object:

e.g., 在屏幕上 MX6 将会加入黑科技
zài píngmù shàng MX6 jiānghuì jiārù hēi-kējì
(“MX6 will add black technology on the screen.”)

iii. Attributive:

e.g., 黑天鹅事件 hēi tiānē shìjiàn (“black swan incident”)

iv. Adverb: no instance

6.2.8 Noun phrase: Countability

Whether the NP containing hēi “black” is count or mass.

i. Count: 黑马 hēi-mǎ (“black horse”)

ii. Mass: 黑公关 hēi gōngguān (“black publication relation”)

6.2.9 Noun phrase: Semantic type: dep-relation with hēi “black”: sub

Whether the semantic type of subject collocated with hēi “black” is of type:

i. Abstract entity: 宣传 xuānchuán (“propaganda”)

ii. Body part: 心 xīn (“heart”)

iii. Animate: 张馨予 Zhāng-xīnyū (“NAME: Zhang Xinyu”)

iv. Inanimate object: 信 xīn (“letter”)

49
v. Organization: 物业公司 wùyè gōngsī ("property management company")

6.2.10 Noun phrase: Semantic type: dep-relation with: dobj
Whether the semantic type of direct object collocated with  hēi “black” is of type:

i. Abstract entity: 信息 xìnxī ("information")

ii. Body part: 脸 liǎn ("face")

iii. Animate: 黄子韬 Huáng-zītāo ("NAME: Huang Zitao")

iv. Inanimate object: 鸟巢 Niǎocháo ("Bird's Nest")

v. Organization: 华为 Huáwèi ("Huawei")

6.2.11 Noun phrase: Semantic type: dep-relation with hēi “black”: iobj
Whether the semantic type of indirect object collocated with  hēi “black” is of type: Abstract entity, Body part, Animate, Inanimate object, Organization. (Notes: No instance in the dataset.)

6.2.12 Verb phrase: Semantic type: Noun collocated with hēi “black”
Whether the semantic type of noun collocated with hēi “black” in a VP is of type:

i. Abstract entity: no instance

ii. Body part: (黑着)脸 (hēi zhe) liǎn ("face blackened")

iii. Animate: no instance
iv. Inanimate object: no instance

v. Organization: no instance

6.2.13 Noun phrase: Semantic type: hēi “black” as attribute
Whether the semantic type of NP where hēi “black” as attribute is of type:

i. Abstract entity: (黑)一日游 (hēi) yīrìyóu (“(black) one day trip”)

ii. Body part: (黑)脸 (hēi) liǎn (“black face”)

iii. Animate: (黑)导游 (hēi) dǎoyóu (“black tour guide”)

iv. Inanimate object: (黑)出租 (hēi) chūzū (“black cab”)

v. Organization: (黑)工厂 (hēi) gōngchǎng (“black factory”)

6.2.14 Noun phrase: Semantic type: hēi “black” as head (including the single hēi “black”)
Whether the semantic type of NP where hēi “black” as head is of type:

i. Abstract entity: 地域黑 diyuē-hēi (“regional discrimination”)

ii. Body part: no instance

iii. Animate: 黑(转粉) hēi zhuǎn fěn (“from detractor to fan”)

iv. Inanimate object: no instance
6.3 Discourse Information
This section illustrates the contextual features focusing on discourse information.

6.3.1 Clause: Clause type
Whether the clause containing 黑 “black” is of type:

i. Main Clause:
   e.g., 她无意中黑了青岛一把。
   tā wúyìzhōng hēi le Qīngdǎo yī-bā
   (“She unintentionally blackened Qingdao one time.”)

ii. Dependent Clause: see 6.3.2

6.3.2 Clause Types of dependent clause
Whether the dependent clause is of type:

i. Adverbial Clause:
   e.g., 他[黑着脸], 急速而有力地嚼着口香糖, 使出全身解数拯救垂死的公牛。
   tā hēi zhe liǎn, jísù ér yǒulìde juézhe kǒuxiāngtáng, shǐchū quánshēnxièshù zhěngjiù chuísǐ-de Gōngniú
   (“With a dark face, he chewed the gum hurriedly and vigorously, trying his best to save the dying Bulls.”)

ii. Relative Clause:
6.3.3 Sentence: The omission of co-arguments

Whether the collocated arguments of 黑 “black” is omitted.

e.g., 萧敬腾经纪公司网页被黑。

萧敬腾经纪公司网页被黑。

6.3.4 Sentence: Pronouns

Whether 黑 “black” collocated with a pronoun.

e.g., 估计你皇兄等到脸都黑了。

估计你皇兄等到脸都黑了。

6.3.5 Sentence: Explication

Whether the meaning of 黑 “black” is explicated in the instance.

e.g., 我买咸鸭蛋也被黑过，[标价 10.80 元/6 只，结账 12.60 元]。

我买咸鸭蛋也被黑过，[标价 10.80 元/6 只，结账 12.60 元]

When I bought salted duck eggs, I have also been hoodwinked by being charged 12.60 yuan for 6, which is priced at 10.80 yuan per 6.”
6.3.6 Pragmatic: Ambiguous

Whether the meaning of  
“black” is ambiguous.

* e.g., 以新秀挑大梁的中国队比韩国队还黑。
  *“The Chinese team, which is a rookie, is even more black than the South Korean team.”* (black can be either interpreted as “Unexpected” or “Malevolent”)

6.3.7 Pragmatic: Mood

Whether the instance is of declarative, interrogative or imperative mood:

i. Declarative:
   * e.g., 这家网站已经被黑。
     *“This website has been hacked.”*

ii. Interrogative:
   * e.g., 为什么被黑?
     *“Why it was slandered?”*

iii. Imperative:
   * e.g., 开网店不要太黑!
     *“Don’t be too black when running a e-shop!”*

7. Appendix 2
For the readers who don’t know Chinese, this section provides a brief introduction for the functional words occurred in the cited Chinese examples.

**Table 6. Introduction for the functional words**

<table>
<thead>
<tr>
<th>Functional Words</th>
<th>TAG</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>是</td>
<td>SHI</td>
<td>Be; Yes</td>
</tr>
<tr>
<td>了</td>
<td>LE</td>
<td>Aspectual Marker for temporally anchored and realized event</td>
</tr>
<tr>
<td>的</td>
<td>DE₁</td>
<td>Possession or Modification Marker</td>
</tr>
<tr>
<td>被</td>
<td>BEI</td>
<td>Passive Marker</td>
</tr>
<tr>
<td>得</td>
<td>DE₂</td>
<td>Degree Complement</td>
</tr>
<tr>
<td>不</td>
<td>BU</td>
<td>Negation Marker</td>
</tr>
<tr>
<td>着</td>
<td>ZHE</td>
<td>Continuous Aspect Marker</td>
</tr>
<tr>
<td>项; 笔; 匹</td>
<td>CL</td>
<td>Classifier</td>
</tr>
<tr>
<td>吗</td>
<td>MA</td>
<td>Question Particle</td>
</tr>
</tbody>
</table>