ARTICLE

Check for updates

Entity, event, and sensory modalities: An onto-cognitive account of sensory nouns

Yin Zhong [1], Kathleen Ahrens[2] & Chu-Ren Huang [3✉]

Nouns in human languages mostly profile concrete and abstract entities. But how much eventive information can be found in nouns? Will such eventive information found in sensory nouns have anything to do with the cognitive representation of the basic human senses? Importantly, is there any ontological and/or cognitive motivation that can account for this noun–verb dichotomy via body-and-world interactions? This study presents the first comprehensive investigation of sensory nouns in Mandarin Chinese, examining their qualia structures formalised in the Generative Lexicon Theory, as well as the time-dependent (endurant–perdurant) properties encoded in their sensory modalities. This study fills the gap in sensorial studies by highlighting the pivotal position of nouns in sensory experiences and provides insights into the interactions between perception, cognition, and language. Further, it establishes, for the first time, the cognitive motivation of the categorial noun–verb bifurcation without presupposing any a priori knowledge of grammatical categories.

[1] Center for Language Education, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR. [2] Department of English and Communication, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR. [3] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR. ✉email: churen.huang@polyu.edu.hk

Introduction

The nature of grammatical categories (i.e. parts-of-speech; PoS), especially the two binary features that account for the major PoSs (i.e. [±N] and [±V] or noun-verb bifurcation), has been a topic of debate at various stages in the history of scholarly pursuits and over a wide range of disciplines. The debate can be traced back to the Greek philosophical discourse on ontology and epistemology (e.g. Plato's *Sophist* and Aristotle's *The Interpretation*) and remains a foundational issue in different linguistic theories. More recently, it has been one of the central issues in the neuro-cognitive studies of human conceptualisation (Gentner and France, 1988), as well as computational approaches to knowledge representation and processing (Redington et al., 1995; Redington et al., 1998). Given the historical depth and disciplinary breadth, it is foreseeable that the terms used and their definitions may be confusing or misleading when framing an interdisciplinary discussion. Hence, we start with the issues to be addressed. This common ground will be built on the philosophical distinction of contrasts between ontological and epistemological beings, as well as the recent theory of embodied cognition that has emerged as a tenet of the theories of mind and cognition in the new century (cf. Clark, 1997; Wilson, 2002).

To begin, and perhaps to risk oversimplification, we view ontology as the study of "beings," i.e. the "nature" of existence, and epistemology as the study of how to identify, verify, or represent the nature of these beings. In contrast, for embodied cognition, the "embodied" refers to the existing physical world, and the "cognition" refers to how human beings, based on their interaction with the outside world, construe and represent in their minds a system of knowledge that is sharable with other human beings. In the context of the theoretical premises given above, the study of language provides a unique opportunity to bridge these two foundations of scientific knowledge. Human languages are viewed as knowledge systems shared by their speakers for sharing and integrating knowledge (Huang et al., 2010a). However, given the vast diversity of human languages, especially in terms of the differences in how nouns and verbs are encoded, our study of the categorial noun–verb bifurcation will not focus on the roles and functions of nouns and verbs in the linguistic systems or human cognitive processing. In addition, given the meta-theoretical dilemma that the nature of beings cannot be discussed without invoking the mapping from beings to linguistic representations, it is also not feasible to tackle the nature of beings with language as the primary data. Instead, we will focus on the following less explored questions that are both intriguing and not directly constrained by the data issues discussed above: what is the basic ontological concept, with null or minimal prior knowledge, that allows for the conceptualisation of shared human experiences to form a foundation of human languages? More specifically, in terms of the system of grammatical categories in human languages, does this shared foundation of conceptualisation lead to a noun–verb bifurcation, creating two of the most basic grammatical categories? Alternatively, in terms of embodied cognition, what is the fundamental characteristic of the physical world that allows human beings, without a priori concepts, to form a shared principle of conceptualisation on which to build the basic noun vs. verb categorial bifurcation? In general, we are looking for the most basic, and ideally, only one, ontological concept that can be experienced and/or perceived without prior knowledge that would, in turn, facilitate aspects of human cognition, such as those reflected in the shared categorial system of human languages.

The experience and knowledge of the interactions between human bodies and their environments are typically acquired through the five sensory modalities: namely, the *visual*, *auditory*, *gustatory*, *olfactory*, and *tactile* senses. These sensory modalities receive the five sensations respectively, i.e. *vision*, *hearing*, *taste*, *smell*, and *touch*. These five senses, also known as the Aristotelian senses, are the conventional categories of fundamental human perception in the literature of various sensorial studies (e.g. Lynott and Connell, 2013; Zhao et al., 2019). To explore how human beings conceptualise sensory experiences of the outside world through languages, the sensory lexicon is both a repository of sensory input, organised and encompassing main lexical categories in human languages, including verbs such as *look* and *hear*, adjectives like *red* and *sweet*, and nouns exemplified by *sight* and *sound*. One interesting fact is that lexical categories may have different tendencies toward particular sensory modalities (Strik Lievers and Winter, 2018). For example, verbs were found over-represented for the auditory and tactile senses but under-represented for vision, whereas adjectives opt for the visual and olfactory senses but less favoured touch and sound in the repertoire of sensory words in English. Such tendencies could further differentiate and reflect the nature of human senses; for instance, the verb-inclination feature carried by the auditory lexicon reveals the dynamic nature of sound. Therefore, the close relationship between lexical categories and sensory modalities is believed to provide empirical evidence in exploring the cognitive basis of grammatical categories using sensory lexicon as the dataset.

Apart from attempting to tackle the foundational issue of conceptualisation by looking at the cognitive motivation of the primary noun–verb dichotomy in human language and through the body-and-world interaction as mediated by sensory perception, it should also be clear that we also need a robust and versatile framework of meaning representation that would be felicitous in terms of cognitive studies, linguistics, and ontology. Aristotle's qualia structure, as the representation of experiential knowledge, presents itself as a natural candidate. In order to be better equipped to deal with modern theories of cognitive sciences and formal ontology, we will adopt the updated version of the qualia structure as proposed and formalised by Pustejovsky (1991, 1995) in the Generative Lexicon Theory (henceforth, the GL theory, or GL). The GL theory, through its inclusion of telicity and agentivity in the qualia structure following Aristotle, provides a rigorous and empirically sound model for the encoding of eventive information by nominals. In addition, the substantial literature on GL-based research on the Chinese language can support the current study (e.g. Song and Huang, 2018).

This paper presents a comprehensive investigation of sensory nouns, given that the sensory lexicon is the repository of integrated human perceptual information. Thus, these nouns are the results of the conceptualisation of relatively well-defined physical contacts of human beings with the physical world. By exploring the correlation between the physical world and sensory concepts, this study aims to explicate the cognitive foundation of grammatical categories. One important precedent is Strik Lievers and Winter's (2018) study of the English sensory lexicon, which found evidence for the cognitive representation of "verby" and suggested that this was the cognitive foundation of the noun–verb categorisation. We seek to substantiate and explicate this proposed cognitive motivation in order to establish the conceptual and/or ontological motivation for the encoding of grammatical categories in other languages, such as Chinese. We will argue that it is the ontological spatio-temporal continuum (and not the nature of an entity) that provides the most robust accounting of the classic concrete–abstract dichotomy.

Theoretical constructs

Grammatical categories: based on the feature system.
In this section, we first lay a foundation for our study by reviewing the

standard linguistic theory of PoSs. We first review the two binary features that account for the four major PoSs (Noun, Verb, Adjective, and Adverb) in formal generative theories, [±N] and [±V] (e.g. Chomsky, 1970). These two features are intuitively defined as being "nouny" (i.e. noun-like) and being "verby" (i.e. verb-like), respectively, so as to account for the feature value assignments of nouns as [+N, −V] and verbs [−N, +V]. Given the four major categories, the other two combinations are assigned as adjectives [+N, +V] and adverbs [−N, −V] (e.g. Baker, 2003; Haegeman, 1994). However, following the intuitive "verby" and "nouny" definitions of the two features, the two assignments for adjectives and adverbs are not clearly justified. In addition, for the same rationale, should not both the deverbal nouns (nouns derived from verbs) and the denominal verbs (verbs formed from nouns) be the most natural candidates for the [+N, +V] assignment as they are attested to have both noun-like and verb-like behaviours? The fact that they are assigned the feature according to their derived categories, e.g. [+N] for deverbal nouns and [+V] for denominal verbs, confirms the tautological nature of the intuitive definition. This widely accepted and practised feature system in linguistics underlines one of the common dilemmas of previous attempts to define grammatical categories by assuming a priori knowledge of them.

Given this circularity, the question then becomes: can grammatical categories be learned without the knowledge of any grammatical categories? The answer is a definite yes. Redington et al. (1995, 1998) demonstrated not only that PoSs can be automatically learned from an untagged corpus (Redington et al., 1995) but also that there are psychologically feasible mechanisms to support such learning without prior knowledge (Redington et al., 1998). Most current computational language processing systems with powerful machine-learning algorithms also attested to the plausibility of learning many different grammatical categories. Contrary to the widely held assumption that these powerful learning mechanisms based solely on distributional information are not interpretable, Chersoni et al. (2021) showed that what is learned by automatic machine-learning mechanisms can be interpreted in terms of semantic features. These studies showed that linguistic categories can be learned (from distributional information) without a priori concepts of the category and that it is plausible to account for such categorical learning to learn more basic features. Taking the lead from these studies, we explore the conceptual a priori's that could lead to the conceptualisation of grammatical categories, in order to overcome the common tautological flaws in previous attempts so as to provide a clear account of the nature of grammatical categories.

The challenge, of course, remains to be the identification of such conceptual a priori's, without any currently held linguistic knowledge. There are, in fact, several potentially viable and conceptually related proposals to account for the binary noun–verb bifurcation. Givón (2001), for instance, argued that nouns could be identified by their "temporal stability." Gentner (1982) and Ahrens (1999), among others, showed that verbs are more mutable (i.e. have a higher propensity to change) than nouns. More recently, Strik Lievers and Winter (2018), as mentioned above, suggested that "eventivity" is the cognitive motivation for verbs. Among these proposals, of special interest to the current study, is Aristotle's elegant definition that "…a noun…has no reference to time" and "a verb…carries with it the notion of time."[1] Aristotle's definition of with or without the notion of time has more recently been adopted in several formal ontologies as the foundation of knowledge systems, such as DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering; Gangemi et al., 2002) and BFO (The Basic Formal Ontology; Arp et al., 2015). In these ontologies, all entities are construed either as enduring continuants, i.e. that "existed in time

and have no temporal parts," or as perduring occurrents, i.e. that consist of "temporal parts and [that] have phases and temporal slices corresponding to the intervals and moments through which [the parts] perdure" (Simons and Melia, 2000, pp. 59–60).

Note that a fundamental challenge to the theories of conceptualisation or categorisation is required a priori knowledge to form a concept or to define a category. Thus the question we pose is: given the ultimate *tabula rasa* of a single spatio-temporal continuum, what is the minimal prior knowledge needed to support conceptualisation? The endurant–perdurant bifurcation requires the knowledge of "reference to time," which may or may not be inherent to the nature of the spatio-temporal continuum. In contrast, concrete–abstract or entity–eventivity bifurcations require a priori knowledge of two categories and likely additional knowledge of how they differ. Similarly with even higher requirements, the [±N] and [±V] theories of the grammatical system are built on prior knowledge of what nouns and verbs are like. By Occam's razor, "reference to time" is the minimal premise required given the *tabula rasa* of the spatio-temporal continuum, and hence our null hypothesis until proven otherwise. Therefore, we seek evidence to support the idea that this foundational concept of reference to time also underlies the formation of a categorial system in human languages.

In what follows, we will delineate the problems of using an entity–eventivity dichotomy to represent the noun–verb bifurcation and further explicate our proposal to adopt time-independent properties, i.e. an endurant–perdurant dichotomy, so as to capture the knowledge representation manifested in one of the major PoSs that is under examination herein, i.e. the noun.

Classification of nouns: based on the entity–eventivity dichotomy. It is commonly believed that PoSs have cognitive origins (e.g. Gentner, 1978; Langacker, 1987). Prototypical verbs mostly represent events and processes, while prototypical nouns are concrete, abstract, or imaginary entities (e.g. Langacker, 1987). However, although entities are typically encoded as nouns while events are surfaced as verbs, not all "verbs" represent events, much like not all "nouns" signify entities.[2] One set of linguistic facts that could shed light on the nature of noun-verb bifurcation is the deverbal nominals. These are nouns derived from verbs and thus retain the notion of eventivity. In particular, we take data from Chinese, a language with minimal morphological marking of inflection and derivation (Hsieh et al., 2022), as there are typically no overt markers to differentiate derived and non-derived forms, so as to provide a unique data set and associated opportunity to examine the nature of the ontological changes, independent of the effect of morpho-lexical rules. For example, we show in (1)–(4) *bàogào* 'report; to report', which has the original lexical category of a verb (1), can also function as a process nominal (2), a result nominal (3), and an event nominal (4) with the same wordform.

(1)	Tā	bàogào	le		sān	ge	xiǎoshí.[4]
	s/he	report	ASP[3]		three	CL	hour
	'S/he reported for three hours.'						
(2)	Tā	zhèngzài	zuò	bàogào.			
	s/he	currently	make	report			
	'S/he is (making) reporting.'						
(3)	Tā	jiāo	le		yī	fèn	bàogào.
	s/he	submit	ASP		one	CL	report
	'S/he submitted a report.'						
(4)	Zhè	chǎng	bàogào(huì)[5]	chíxù	le	sān	ge	xiǎoshí.
	this	CL	report(meeting)	continue	ASP	three	CL	hour
	'This report (meeting) lasted for three hours.'							

The differences between "process nominal" and "result nominal" are well established in past literature on nominalisation

Table 1 Classification of nouns based on entity–eventivity dichotomy.		
Type of nouns		**Examples**
Entity	Nouns referring to objects	*shù* 'tree,' *zhuōzi* 'table,' *huǒchēzhàn* 'train station'
	Deverbal nominals/ Deverbal nouns: result nominals	*bàogào* 'report,' *fānyì* 'translation,' *yánjiū* 'research'
Eventivity	Deverbal nominals/Deverbal nouns: process nominals	*bàogào* 'report,' *fānyì* 'translation,' *yánjiū* 'research'
	Nouns referring to events	*huìyì* 'meeting,' *bàoyǔ* 'rainstorm,' *zǎofàn* 'breakfast'

(cf. Fu, 1994; Grimshaw, 1990). They can often be differentiated by derivational affixes in many languages, such as English (e.g. *giving* versus *gift*), but not always (e.g. *building* a *building*). In Chinese, process nominals and result nominals select different types of classifiers (Huang and Ahrens, 2003). For instance, *fèn* 'a copy (for documents, newspapers, periodicals)' in sentence (3) is an individual classifier marking *bàogào* in (3) a result nominal. Moreover, *bàogào* can also follow an event classifier *cì* 'occurrence; time' that enumerated events, e.g. *zuò le yī cì bàogào* 'made one time/instance of reporting.' Another determining factor in differentiating the two is whether the noun phrase allows durative time expressions and time points; for example, *sān ge xiǎoshí* 'three hours' can only go with a process nominal but not with a result nominal.

As for the event nominal (or event nouns) in the sentence (4), they are basically the naming of an event as an entity (Chierchia, 1986). The event nouns can be differentiated from nouns referring to physical objects by several empirical criteria: event nouns which (a) can be selected by event classifiers (e.g. *chǎng* [for sporting or recreational activities], *cì* [for enumerated events], *dùn* [for meals, beatings, scoldings, etc.]), (b) can provide argument structure information when serving as an object of light verbs (e.g. *kāishǐ* 'begin,' *jìxù* 'continue,' *tíngzhǐ* 'stop'), (c) can allow temporal noun suffixation to denote temporal duration orientations (e.g. *qián* 'before,' *hòu* 'after,' *zhōng* 'in the course of'), and (d) can allow durative temporal expressions (e.g. *sān gè xiǎoshí* 'three hours,' *shí tiān* 'ten days'), to name a few (Han, 2010; Shao and Liu, 2001; Wang, 2013). It should be noted that previous literature did not converge on a consensus definition or scope of event nouns in Chinese. For example, Deng's (2021) "event noun" is rather similar to the aforementioned "process nominals." He suggested that "simple event nominals" (Grimshaw, 1990) or "(pure) event nouns" (Wang, 2013) could be considered a sub-category of the "process nominals." Nevertheless, in order to differentiate derived from non-derived eventive nouns, in this paper, we follow Wang (2013) and refer to nouns naming events that are not derived from verbs as "event nouns," while those nouns that are derived from verbs are referred to as "deverbal nominals" or "deverbal nouns."

In general, nouns that make reference to physical objects, deverbal nominals (process nominals and result nominals), and nouns referring to events, can be differentiated based on their denotations of entity and/or eventivity, as shown in Table 1.

As seen from the above summary, the entity–eventivity dichotomy does not align with the noun-verb PoS classification. In other words, nouns encompass entities and events, and considerable noun-verb categorical fluidity exists in Chinese (Kwong and Tsou, 2003a, 2003b). Note that although verbal meanings are more mutable than their nominal counterparts (Ahrens, 1999; Gentner, 1982), the degree of ambiguity of noun-verb dual category words is the same for each category unless the direction of change can be identified. Given that the dichotomy of entity–eventivity cannot reliably predict the noun-verb category assignment, an alternative conceptual dichotomy that does not require the intuitive entity–eventivity bifurcation is thus needed.
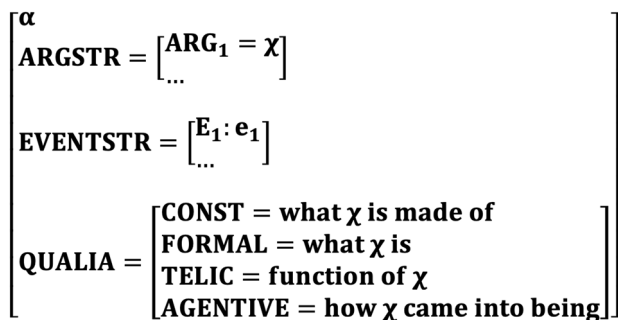
Re-classification of nouns: based on the endurant–perdurant dichotomy. Recall our earlier discussions on the endurant–perdurant bifurcation in formal ontologies, which will serve as the theoretical foundation of the current account. All concepts can be classified according to whether they exist independent of time (i.e. endurant or continuant) or are dependent on time (i.e. perdurant or occurrent). Ontologically, information is described as attributing properties to some constant objects, but for such encoding to work, the "sameness" of the object must be maintained (i.e. its being endurant or continuant); on the contrary, in order to describe changes, there must be properties that can be identified as a variable of time (i.e. being perdurant or occurrent). When such ontological bifurcation is reflected in the language, a noun is simply the default way to encode an endurant concept, as it is a "rigid designator," while verbs, requiring interpretation of their values depending on time, are the default way to encode perdurant concepts (Huang, 2015, 2016).

From this ontological point of view, deverbal nouns and denominal verbs involve type-shifting. Deverbal nouns represent eventive information that is disassociated from temporally bound interpretations. An apt example is the way we refer to a scheduled flight (Huang, 2015). *A 3:30 flight*, with the deverbal noun *flight* derived from the verb *to fly*, does not need to reference any specific point in the spatio-temporal continuum. In other words, a flight is *a 3:30 flight* is not linked to any specific time points in terms of the event. Instead, it is defined by belonging to a class of eventive entities that share the scheduled flight time and can happen at any time. That is, eventive information, such as the time of the scheduled flight, is treated as constant and can be assigned time-sensitive interpretation when co-occurring with a verb, such as *The 3:30 flight took off at 4:30 today*. It is, therefore, possible to use the fundamental conceptual bifurcation of time dependency to conceptualise the linguistic lexical categories to view the entities and their eventive readings from an ontological perspective.

To sum up, ontologically speaking, a noun can be classified as consisting of enduring or perduring features (i.e. maintaining constant referent or not through different times). The current study will focus on sensory nouns in Mandarin Chinese to test the cognitive foundation of grammatical categories differentiated by the endurant–perdurant dichotomy. We will elaborate on the analytical framework (Generative Lexicon Theory) in the next section, followed by data and methodology (corpus-based). Results, as well as the discussion and conclusion, will be provided in the final sections. The main objective of this research is to provide empirical evidence of the linguistic encoding of human sensory experiences as reflected by their heterogeneous nature (endurant or perdurant) within a particular lexical category (the noun) and further shed light on the cognitive foundation of grammatical categories.

Analytical framework
The Generative Lexicon (GL) theory is chosen as the framework for our study based on two important considerations. First, GL has a fully formalised qualia structure that is linked to an

$$\begin{bmatrix} \alpha \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG}_1 = \chi \\ ... \end{bmatrix} \\ \\ \text{EVENTSTR} = \begin{bmatrix} \text{E}_1 : \text{e}_1 \\ ... \end{bmatrix} \\ \\ \text{QUALIA} = \begin{bmatrix} \text{CONST} = \text{what } \chi \text{ is made of} \\ \text{FORMAL} = \text{what } \chi \text{ is} \\ \text{TELIC} = \text{function of } \chi \\ \text{AGENTIVE} = \text{how } \chi \text{ came into being} \end{bmatrix} \end{bmatrix}$$

Fig. 1 Lexical representation of a word (Pustejovsky, 2013, p. 26). Note. α = a word; ARG1 = argument 1; E1 = subevent 1.

Fig. 2 Lexical typing structure with referring to qualia roles.

ontology (Pustejovsky et al., 2006). As such, we can treat it as a fully implemented version of Aristotle's qualia and leverage its theoretical constructs to link to the primitive concept of endurant/perdurant, as well as to represent experiential information of the physical world. Second, a methodologically critical motivation is that GL provides a theory of argument selection based on experiential knowledge (i.e. qualia structure) instead of other theories that rely on grammatical categories or related features. We noted earlier that the failure to fully understand the nature of nouns and verbs in previous studies is probably due to the fact that each of them requires a certain degree of knowledge of grammatical categories. Even the well-designed study reported in Strik Lievers and Winter (2018) relies on prior knowledge of grammatical categories, that is, the PoS assigned by the dictionary and by corpus annotation. Conversely, GL predicts argument realisation in terms of semantic typing (based on qualia information) and by the semantic process of selection, exploitation, and coercion without referring to grammatical categories. The basic structure of GL and its relevance for our current study is explicated below.

In the GL theory, the semantic representation of a word can be represented at four levels: argument structure, event structure, qualia structure, and lexical typing structure[6] (Pustejovsky, 1995, 2013; Pustejovsky and Jezek, 2008), as shown in Fig. 1. Argument structure (ARGSTR) mainly specifies the number and nature of the arguments a nominal phase can take, while event structure (EVENTSTR) identifies the event type and any subeventual structure a word or a phrase may have. Since this study targets nouns, only qualia structure (see the section "Qualia structure") and lexical typing structure (see the section "Lexical typing structure") will be consulted, as these two are more effective in explaining the semantic representations of nouns.

Qualia structure. The Qualia structure in GL is adapted from the Aristotelian qualia with four causes: material cause, formal cause, efficient or moving cause, and final cause (Pustejovsky and Jezek, 2014). As depicted in Fig. 1, the constitutive role (CONST) corresponds to material cause, as it describes the relationship between an entity and its constitutive parts, as well as the relation between the parts and the entire entity by referring to what the entity is made of. The formal role (FORMAL) focuses on how the specific entity distinguishes from other objects within a larger domain; in other words, it encodes taxonomic information and carries information about the basic conceptual category (Pustejovsky and Jezek, 2014). The telic role (TELIC), being the final cause, dealing with the purpose and function of the entity, includes direct telic and indirect telic. The last agentive role (AGENTIVE), corresponding to the efficient or moving cause, involves factors related to the entity's origin that force the entity to come into being. As mentioned earlier, we treat qualia
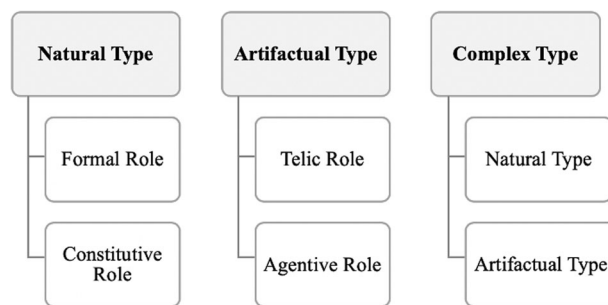
structure as the collection of lexically conventionalised experiential knowledge, following Aristotle's original design.

Lexical typing structure. Pustejovsky (2001, 2013) further divided the domain of individuals into three types based on the four fundamental qualia roles in the qualia structure, as illustrated in Fig. 2.

As shown in Fig. 2, natural types differ from artefactual types in their references to formal and constitutive roles only. *Noise* is an example of a natural type, as its lexical meaning inherits directly from its superordination, and only the formal and constitutive role of *noise* (i.e. *sound*) is exploited in the common use of this word. Conversely, an artefactual type refers to telic and agentive roles, especially emphasising the function or purpose of the object. For example, *piano* is considered an artefact, given that the purpose of a piano lies in its telic role, which is to be played and to allow people to listen to the melody being played. Last but not least, a complex type (or dot object) makes references to the relation between the above two types. *Song* is an instance of a complex type because of its composition of sound and information (i.e. [sound·info]). For instance, in sentence (5), the meaning facet of *sound* is elicited because *melodic* describes the melody of the song, while in (6), information being mainly activated as *inspirational* indicates that the lyrics of the song are encouraging, which conveys information.

(5) This is a melodic song.
(6) This is an inspirational song.

By referring to the qualia structure and the lexical typing structure, we can assign endurant features to those natural types in which only their formal and/or constitutive roles are exploited in their meaning representations. As for artefactual types, since they focus on the function or purpose of the objects and how these objects come into being, they may either be endurant or perdurant, depending on which meaning facet has been selected via type coercion (Pustejovsky, 2001; Pustejovsky and Jezek, 2008). For instance, in the sentence of *I saw a piano*, the endurant feature of the *piano* is selected because the meaning of *piano* in such a visual event refers to a physical object that does not reference time points; however, when an event (mostly due to different verbs) emphasises on the function of the object, such as in *I heard (someone is playing) the piano*, the perdurant, or temporarily bounded properties of the *piano* will take effect in order to participate in the meaning coercion. Complex types, referring to natural and artefactual types, shall be considered perdurant when a time concept is involved. This is because coprediction for these *dot objects* is always allowed, e.g. *This book is long and interesting*. In other words, this type of noun does not need to go through the type-shifting process in order to get its meaning across, and its eventive meaning is always presented. We

Table 2 Relations between lexical typing structures and endurant–perdurant dichotomy.

Lexical typing structures	Natural type	Artefactual type	Complex type
Endurant–perdurant dichotomy	Endurant	Endurant or perdurant	Perdurant

summarise the correspondence between the three lexical typing structures and the endurant–perdurant dichotomy in Table 2.

Method

Data collection. The relevant data will be collected following the motto of corpus linguistics as proposed by Firth (1962) "you shall know a word by the company it keeps." To study sensory nouns, we will start with the words accompanying them, especially the objects selected by the perceptual predicates. Thus, this study partly follows Pustejovsky and Jezek's (2008) corpus-based investigation of identifying mechanisms of semantic coercion in predicate-argument constructions.

To examine the largest possible number of sensory nouns, we used the basic sensory verbs as the predicates of the five sensory modalities to extract the sensory events in the corpus, including visual events (indicated by the visual verbs *kàn* 'to look,' *jiàn* 'to see,' and *kàn/jiàn-dào* 'saw'), auditory events (indicated by the auditory verbs *tīng* 'to listen,' and *tīng-dào* 'heard'), gustatory events (indicated by the gustatory verbs *cháng* 'to taste,' and *cháng-dào* 'tasted'), olfactory events (indicated by the olfactory verbs *wén* and *xiù* 'to smell; to sniff,' and *wén/xiù-dào* 'smelt'), and tactile events (indicated by the tactile verbs *mō* and *chù* 'to touch,' *gǎnjué* 'to feel,' *mō/chù-dào* 'touch; feel,' and *gǎnjué-dào* 'felt'). Note that although *gǎnjué* 'to feel' usually is not considered a typical tactile verb, it is defined as "to perceive and distinguish external stimuli via bodily sensations" in the *Chinese WordNet 2.0* (Huang et al., 2010b);[7] therefore, this word is highly tactile-related and is believed to trigger bodily feelings.

The data extraction and analysis procedures are the following:

a. To extract the set of nouns from the corpus that typically co-occurs with the verb in a specified grammatical relation. For our current purposes, we restrict our investigation to the relation of *object-of* the sensory verbs.
b. To annotate the selected nouns with the foregoing qualia structures.
c. To classify those nouns into three types and analyse their respective characteristics with reference to their qualia values.

All of the data and sentence examples presented in what follows, unless otherwise specified, were extracted from a Chinese online corpus, *Chinese Web 2011 (zhTenTen11)* in *the Sketch Engine* (Kilgarriff et al., 2014).[8]

Data cleaning. Since objects that the sensory modalities can perceive are the primary concerns in this study, we will only look at those nouns that elicit perceptual information. In what follows, we will use visual perception (through the visual verb *kàn* 'to look; see') as an example to demonstrate how we identify and select sensory nouns from the corpus.

First, we use the *Word Sketch* function in the *Sketch Engine* to generate an exhaustive list of objects collocated with the keyword *kàn* 'to look; see.'[9] Next, as sketched in the *Chinese WordNet 2.0*, among the various meanings of *kàn* 'to look; see,' two meanings are associated with visual perception, including 'to perceive

Table 3 Distribution of lexical typing structures in meanings of *kàn* 'to look; see'.

Lexical typing structures	Meanings of *kàn* 'to look; see'		Total number of visual nouns
	To perceive through sight	To understand or to appreciate through sight	
Natural type	124 (77.5%)	0 (0%)	124 (40%)
Artefactual type	27 (16.9%)	**86 (57.3%)**	113 (36.5%)
Complex type	**9 (5.6%)**	**64 (42.7%)**	73 (23.5%)
Total	160 (51.6%)	150 (48.4%)	310 (100%)

Perdurant concepts are shown in bold.

through sight' and 'to understand and appreciate through sight,' as shown in the sentences (7) and (8), respectively:

(7) Nǐ zài qiáo-shàng **kàn fēngjǐng**, kàn fēngjǐng de rén zài lóu-shàng **kàn nǐ**
you on bridge_on look scenery look scenery DE person on building_on look you
'You are enjoying the scenery on the bridge while the people enjoying the scenery are looking at you from upstairs.'

(8) Yuánběn shǔyú dàzhòng yúlè de **kàn diànyǐng** biànchéng le gāo xiāofèi.
before belong public entertainment DE watch movie become ASP high consumption
'Watching movies was previously deemed an entertainment for the general public, but it is an expensive consumption nowadays.'

The above two examples showed that *kàn* 'to look; see' can take objects such as *fēngjǐng* 'scenery' and person (e.g. *nǐ* 'you') in (7), which evokes the meaning of 'to perceive through sight'; *kàn* 'to look; see' also co-occurs with the object like *diànyǐng* 'movie' in (8), with the meaning associated with 'to understand and appreciate through sight.' Therefore, we only consider the objects selected by these two meanings in the visual events. All the extended meanings and/or metaphorical meanings of the sensory verbs are not considered.

Results

Visual nouns. 310 nouns selected by *kàn* 'to look; see' were identified from the corpus after data cleaning. The overall distribution of three lexical typing structures in the two meanings of *kàn* 'to look; see' is presented in Table 3. Note that categorising lexical typing structures is strictly pertinent to the specific sensory domain being examined. For example, *fēngjǐng* 'scenery' is considered a natural type in visual events rather than other possible types in other perceptual events. The time-dependency conceptualisation of each cell is determined by the interaction of the lexical types and the predicate. For easy reference, perdurant terms are shown in bold, in Table 3.

To perceive through sight. Among the 160 nouns that elicit the meaning of "to perceive through sight," natural type is the majority, constituting 77.5% of the three structures. The natural type exploits the formal and constitutive roles of a noun; in most circumstances, the formal role plays a key role in the collocative meaning between *kàn* 'to look; see' and its objects. After attributing ontological categories to the visual nouns related to the meaning of "to perceive through sight," the formal roles of the natural types mainly fall into *appearance* (e.g. *yàngzi* 'appearance; look'), *colour value* (e.g. *báisè* 'white'), *lights* (e.g. *guāngxiàn* 'light'), *location* (e.g. *zhōuwéi* 'surrounding'), *natural things* (e.g. *tiānkōng* 'sky') and *scene* (e.g. *fēngjǐng* 'scenery') based on their ontological taxonomies. Apart from a large number of natural types, 16.9% of the nouns are labelled as artefactual types. This

Table 4 Distribution of lexical typing structures in meanings of *tīng* 'to listen; hear'.

Lexical typing structures	Meanings of *tīng* 'to listen; hear'			Total number of auditory nouns
	To perceive sound through hearing	To appreciate (sound) through hearing	To receive information through hearing	
Natural type	21 (21%)	0 (0%)	0 (0%)	21 (5.5%)
Artefactual type	0 (0%)	**17 (17%)**	**88 (47.6%)**	105 (27.3%)
Complex type	**79 (79%)**	**83 (83%)**	**97 (52.4%)**	259 (67.3%)
Total	100 (26%)	100 (26%)	185 (48.1%)	385 (100%)

Perdurant concepts are shown in bold.

group primarily contains *images* (e.g. *túpiàn* 'picture') and *artefacts* (e.g. *yānhuā* 'fireworks'). Note that the complex type (5.6%) is not salient in the visual events related to "to perceive through sight." Since these are *dot objects*, they refer to the physical objects and the eventive information it carries. For example, a *shǒubiǎo* 'watch' is made for a particular purpose, that is, for people to check the time. Hence, "taking a look at a watch" necessarily involves perceiving the physical watch, as well as telling the time, which refers to the telic role of a *watch*.

To understand or to appreciate through sight. The second meaning of *kàn* 'to look; see' is "to understand or to appreciate the content of or information about the object being looked at." Since this type of object, in most cases, contains a specific function and is artificially created rather than naturally existing, the artefactual type (57.3%) and, especially, complex type (42.7%) far outweigh the natural type (0%). The artefactual type in the meaning of "to understand through sight" mainly consists of *texts and writings*, such as *wénzhāng* 'article' and *xīnwén* 'news,' while *entertainment-related* items such as *diànyǐng* 'movie' and *jiémù* 'programme' give rise to the meaning of "to appreciate through sight." The main difference here lies in the purpose of the entity—the former is created to meet information needs, whereas the latter is mostly used to satisfy entertainment or leisure demands.

Complex type comprises words that have *physical references* but also maintain *information* or *entertainment* functions. Some examples include *shū* 'book' and *bàozhǐ* 'newspaper' (to understand), as well as *diànshì* 'television' and *zhǎnlǎn* 'exhibition' (to appreciate). Also, note that both the physical entity (9) and the information carried in the entity (10) can be exploited in the visual events, as exemplified in the following two examples:

(9)	Wǒ	kàn	zhe	shūguì	shàng	lín-láng-mǎn-mù		de	shū…
	I	look	ASP	bookshelf	on	a_dazzling_array_of_beautiful_things		DE	book…
	'When I was looking at the dazzling array of books on the bookshelf….'								
(10)	Nǐ	zhēn	shi	gè	ài	kàn-shū		de	háizi!
	you	really	be	CL	love	read		DE	child
	'You are a child who likes reading!'								

In the meanings related to "to appreciate through sight," there exist a few words that not only contain *information* but also involve *events and affairs* ([event·info]), including *yǎnchànghuì* 'concert,' *zúqiúsài* 'football game,' *chēzhǎn* 'motor show,' etc. This type of noun is categorised as event nouns. Although the number of event nouns is not very productive in visual nouns, that existence hints that sensory nouns may also denote eventive information to a certain extent.

Summary. In sum, we have shown above that the visual verb *kàn* 'to see' has two senses: one for perception and one for integration of visual information (understanding or appreciating using cognitive skills). The most frequently attested instances are the perception of natural types, which can be considered prototypical visual cognition. Interestingly, the ratio of natural type versus

artefactual type is only roughly 11/10 (124/113). Overall, the distribution suggests that vision is a dominant and versatile sensory domain.

Auditory nouns. As sketched in the *Chinese WordNet 2.0*, apart from its original meaning, "to perceive sound through hearing" (e.g. *tīng shēngyīn* 'listen to the sound'), the semantic facets *tīng* 'to listen; hear' also denote "to appreciate (sound) through hearing" (e.g. *tīng yīnyuè* 'listen to music') and "to receive information through hearing" (e.g. *tīng yǎnjiǎng* 'listen to the speech'). After data cleaning, a total of 385 nouns related to the above three meanings of *tīng* 'to listen; hear' were collected. Table 4 presents an overview of the distributions of three lexical types in the three meanings of *tīng* 'to listen; hear.'

To perceive sound through hearing. In the natural type that evokes the meaning "to perceive sound through hearing," the words mainly relate to different *aspects and physical qualities of sound* (e.g. *shēngyīn* 'sound,' *yīnliàng* 'volume,' *yīnzhì* 'sound quality,' *jiézòu* 'rhythm'). However, the majority of the perceived sound objects fall into the complex type (79%), given that all the words under this category are events concerning both the facets of *sound and event* ([sound·event]). We further identified three primary categories in this group. First, nouns that are induced by events, and they are primarily compound nouns, including *qínshēng* 'the sound of playing instruments,' *gēshēng* 'the sound of singing,' *jiǎobù-shēng* 'the sound of footsteps.' The second type is (simple) event nouns, e.g. *fēng* 'wind,' *yǔ* 'rain,' *hǎilàng* 'waves,' *liúshuǐ* 'running water.' The third type is deverbal nouns, including *xuānxiāo* 'shouting,' *hūxī* 'breathing,' and *xīntiào* 'heartbeat.'

To appreciate (sound) through hearing. The second meaning of "to appreciate (sound) through hearing" takes an artefactual type that is typically part of *music*, such as *yīnyuè* 'music' and *xuánlǜ* 'melody,' in which the telic role of being listened to is exploited. The number of complex types is much greater than the number of other types (83%), which can be further divided into facets denoting *sound and information* ([sound·info], e.g. *gēqǔ* 'song'), *physical object and sound* ([object·sound], e.g. *yuèqì* 'musical instrument'), *human being and sound* ([human·sound], e.g. *Bèiduōfēn* 'Beethoven'), and *event and sound* ([event·sound], e.g. *yǎnchàng* 'sing'). The type of [sound·info] not only comprises sounds but also incorporates content and information that allow listeners to appreciate the melody as well as the content that the melody holds. In the second type [object·sound], hearing events select this type by exploiting the sound facet and mainly use its telic role to generate auditory-related meaning. While in the third type [human·sound], although they are human beings, *tīng* 'to listen; hear' can resort to their telic roles of singing and performing arts or the agentive roles of writing and composing music.[10] The last type ([event·sound]), on the contrary, contains

Table 5 Distribution of lexical typing structures in the meaning of *cháng* 'to taste'.

Lexical typing structures	Meaning of *cháng* 'to taste'
Natural type	8 (19%)
Artefactual type	34 (81%)
Complex type	0 (0%)
Total	42

Table 6 Distribution of lexical typing structures in the meaning of *wén/xiù* 'to smell; to sniff'.

Lexical typing structures	Meaning of *wén/xiù* 'to smell; to sniff'
Natural type	52 (100%)
Artefactual type	0 (0%)
Complex type	0 (0%)
Total	52

event nouns such as *yǎnchànghuì* 'vocal concert' and *yīnyuèhuì* 'musical concert.'

To receive information through hearing. The last meaning examined is "to receive information through hearing" by virtue of the activity *tīng* 'to listen; hear.' In a similar vein, the nouns are mainly categorised into two types, i.e. artefact (47.6%) and complex (52.4%). The telic role, including to speak, to listen, and to communicate, of the artefactual type is mainly exploited. As for the complex type, the major category involves *information and events* ([event·info]) since the facet of sound is less prevalent here. Two types of nouns are also shown in this category, namely, event nouns (e.g. *kè* 'class,' *jiǎngzuò* 'seminar') and deverbal nouns (e.g. *liáotiān* 'chatting,' *tánhuà* 'talking,' and *huìbào* 'reporting').

Summary. As shown above, the hearing verb *tīng* 'to listen/hear' has three senses: one for perception and two for integration of auditory information. The most frequently attested instances are the perception and integration of complex types. This suggests that hearing involves strong integration of physical and abstract information. This is expected as the perception of music and speech as described requires either explicit or implicit knowledge of systems of abstract concepts such as loudness, melody, pitch, prosody as well as phoneme and tone, based on the physical properties of amplitude, articulation, duration, frequency, etc. Another significant feature is the very low percentage of natural types as the target of perception (5.5%). Overall, the distribution suggests that hearing is a sensory domain that is crucial to the integration of information, especially eventive information.

Gustatory nouns. Adopting the same method, the objects selected by the gustatory verb, *cháng* 'to taste,' are scarce compared to the nouns of the above two sensory modalities. The reason may lie in the single meaning for *cháng* 'to taste,' which is only "to distinguish or taste the flavour of food." Of all the 42 gustatory nouns, no complex type was found, and the artefactual type (81%) was more prevalent than the natural type (19%), as shown in Table 5. Natural type is mainly constituted by the *attribute or attribute values of the flavour of the food*, for example, *zīwèi* ' flavour,' *fēngwèi* 'flavour,' and *xiāngwèi* 'fragrance;', whereas the artefactual type consists of *food that is made to nourish the body or satisfy the appetite*, e.g. *càiyáo* 'dishes,' *měijiǔ* 'fine wine,' and *xiǎochī* 'snacks.'

In sum, the gustatory verb *cháng* 'to taste' has one single sense. There are no attested instances of the perception of complex types. Moreover, the ratio of artefactual type over natural type is roughly 4 to 1 (34/8). On the one hand, this seems unusual in the context of taste involving typically embodied objects. On the other hand, this should be expected as most of the food we ingest is artefactual in the sense of being processed. Overall, the distribution suggests that taste is a sensory domain that has evolved to interact primarily with man-made/packaged ingestible objects and only rarely with the natural environment (i.e. such as via personal farming) (Table 6).

Olfactory nouns. Since two olfactory verbs, i.e. *wén* and *xiù* 'to smell; to sniff,' are commonly used to depict olfactory experiences, nouns collocated to both verbs were examined. 52 olfactory objects were generated, of which all of them being categorised as natural types. *Odour and odour values* are the most common components of olfactory nouns. The most distinctive feature of these nouns is that they are either composed of the morphemes *wèi* 'taste; smell' or *xiāng* 'fragrance' (e.g. *qìwèi* 'odour,' *chòuwèi* 'bad smell,' *fāngxiāng* 'fragrance,' and *qīngxiāng* 'faint scent'). The pattern and structure of the artefactual type are also fairly consistent. They mainly consist of compounds with *wèi* 'taste; smell' and *xiāng* 'fragrance' as the stems and are used to describe the *smell or fragrance of the artefacts*. Examples include *yóuyānwèi* 'the smell of fuel fume,' *fǔchòuwèi* 'a rancid smell,' *jiǔxiāng* 'aroma of wine,' *fànxiāng* 'rice fragrance,' to name a few.

Similar to the gustatory category, the olfactory verb *cháng* 'to taste' has one single sense. The only attested instances involve the perception of natural types; given the low frequency and the strong connection between the gustatory and olfactory senses, the lack of artefactual type is of interest. This may be the result of insufficient data or might be due to the fact that smell, unlike taste, is often not volitional.

Tactile nouns. Finally, tactile nouns were collected by examining the collocations with two tactile-related verbs, i.e. *mō* 'to touch' and *gǎnjué* 'to feel.' In the *Chinese WordNet 2.0*, *mō* 'to touch' is illustrated as "to use hands to touch the object" while *gǎnjué* 'to feel' is "to perceive and distinguish external stimuli via bodily sensations;" hence, two distinct categories of tactile nouns are expected because of the distinct meanings of the tactile predicates.

As presented in Table 7, a total of 58 tactile nouns were generated. It is found that the number related to *mō* 'to touch' (79.3%) far outweighed that collocated to *gǎnjué* 'to feel' (20.7%). The natural type related to the meaning of *mō* 'to touch' embraced *body parts* (e.g. *nǎodai* 'head,' *dùzi* 'belly'), *body substance* (e.g. *jīfū* 'skin,' *pífū* 'skin'), and *salient substance on the body* (e.g. *yìngwù* 'hard substance,' *zhǒngkuài* 'lump'). In the natural type for the meaning of *gǎnjué* 'to feel,' it mainly consists of temperature-related items, especially more abstract experiences such as *nuǎnyì* 'warmth' and *liángyì* 'coolness.' The artefactual type is mostly associated with *mō* 'to touch.' It is primarily comprised of *technology products*, such as *píngmù* 'screen' and *jiànpán* 'keyboard.' In summary, the results showed that nouns selected by *mō* 'to touch' are more related to the tactile perception over the *physical objects*, whereas nouns collocated with *gǎnjué* 'to feel' more indicate *bodily feelings*.

In sum, Table 7 shows two tactile verbs: one *mō* for perception and one *gǎnjué* for integration of tactile information. The touch nouns are dominated by natural types. This is expected as tactile sense is commonly considered the most embodied sense modality. Compared with vision, another sensory domain with a high frequency of natural-type objects, touch has a significant portion of artefactual-type objects and is also much lower in total numbers. Overall, this suggests that the tactile sense is well-grounded but less versatile.

Table 7 Distribution of lexical typing structures in the meanings of *mō* 'to touch' and *gǎnjué* 'to feel'.			
Lexical typing structures	**Meanings of *mō* 'to touch' and *gǎnjué* 'to feel'**		**Total number of tactile nouns**
	To touch	**To feel**	
Natural type	36 (78.3%)	11 (91.7%)	47 (81%)
Artefactual type	10 (21.7%)	**0 (0%)**	10 (17.2%)
Complex type	**0 (0%)**	1 (8.3%)	1 (1.7%)
Total	46 (79.3%)	12 (20.7%)	58 (100%)
Perdurant concepts are shown in bold.			

Table 8 Distribution of endurant–perdurant types for sensory nouns in three sensory modalities.			
Endurant–perdurant types	**Sensory modalities**		
	Vision	**Hearing**	**Touch**
Endurant	151 (48.7%)	21 (5.5%)	57 (98.3%)
Perdurant	159 (51.3%)	364 (94.5%)	1 (1.7%)
Total	310	385	58

Discussion

This section synthesise the above results of sensory nouns according to their involvement of endurant and/or perdurant properties. Note that in the GL theory, natural type objects are concrete entities (i.e. formal and constitutive), thus, they are endurant by nature. Artefactual type objects (i.e. agentive and telic) can be either endurant or perdurant, depending on which meaning facet is coerced by the predicates. In perceptual events, we suggest that endurant properties will mostly be elicited when these artefacts are selected by pure perceptual events (e.g. the objects that are perceived through sight, hearing, etc.); whereas when integration of perceptual information is needed (e.g. objects that are appreciated or understood through sight, hearing, etc.), perdurant features will be resorted to. Taking *piano* as an example again, as demonstrated above, in a visual event, *piano* is considered an artefact that is only perceived via pure perception (i.e. objects that are perceived through sight only); it should be a concrete object that does not involve any time points. In contrast, if the sound of (someone is playing) the piano is appreciated as an auditory event, then *piano* shall be bounded by temporal features because the sound of playing contains temporal intervals, even though piano per se is an object. As for the complex types, they are a combination of entities and events. Thus, we can assume that the complex type following perceptual verbs would be dominated by their natural type meaning, while the same dot objects following information integration type verbs would be dominated by their artefactual type meaning. Obviously, the nature of complex type objects means that the two aspects of their meaning are always accessible regardless of the context. However, it is also reasonable to assume that the selection of the verb will make one aspect comparatively more accessible.

Generally, we mark all the natural types and artefactual types perceived via pure perception to contain endurant features, while artefactual types perceived via integration of perceptual information and all the complex types to contain perdurant features. Based on the results in the above section (Tables 3–7), we further summarise the sensory nouns' involvement of endurant or perdurant properties in Table 8.

From Table 8, we can see that tactile perception is strongly preferred by endurant/time-independent objects, while auditory perception strongly prefers perdurant/time-dependent objects. Lastly, visual perception is most versatile in usage, given that the sense has no strong preference or inclination for either endurant or perdurant properties. Note that we only take sensory nouns as a target of perception and use the lexical typing theory of GL to examine the distribution of sensory objects. Both steps presuppose no knowledge of grammatical categories. Interestingly, note that by combining the above results with our proposal that endurant entities are linguistically represented as nominal units, and perdurant entities as verbs, the current results imply that the categorical assignment of the tactile sensory properties is more likely to instantiate as nominal elements. The strong perdurant dominance means that the auditory sensory properties are more likely to be expressed as verbal elements. The above results corroborate the results from the category counting study of Strik Lievers and Winter's (2018) on English. Recall that they showed that touch is over-represented by nouns, and hearing is over-represented by verbs. They also report an over-representation of vision by adjectives, which was not accounted for. Our results, however, can be interpreted as vision's allowing it to both modify nominal category adjectives and serve as a source domain in linguistic synaesthesia (cf. Zhao et al., 2019), which also favours vision occurring in pre-nominal modifying positions.

Note that we only list three sensory domains in the above table. There are three reasons to omit olfactory and gustatory senses in this comparison. The first is the relative sparseness of data, as observed above. The second is that the lack of an information integration verb in the data for these two sensory domains makes it impossible to perform a reliable direct comparison. Third, we observe that neuro-cognitive studies more typically involve the three senses of vision, hearing, and touch (using the more updated term of somatosensory). For instance, Sanchez et al. (2020) compared these three senses in terms of late latency and interpreted the results in terms of conscious perception. We compare our findings to their findings which used a different methodology below.

The comparative study of three sensory modalities based on MEG measurement of brain activities reported by Sanchez et al. (2020) has some interesting parallelism with our results. Their study aimed to establish a supramodal brain network for processing all senses, as well as to differentiate these three senses, given that they share the same supramodal network for processing. After establishing the shared uses of a supramodal network, Sanchez et al. (2020) showed significant differences in late latency among their senses. In particular, they showed that hearing and vision have later latency overlapping with the P300 area, and the somatosensory system has relatively earlier latency than the other two. The proposed account of this difference is based on the assumption that the somatosensory does not involve conscious perception, while vision and hearing do. Conscious perception means more efforts are needed to integrate and represent sensory information instead of a simple recording of sensory data from a non-conscious perception. Sanchez et al.'s (2020) results are compatible with the result of our study of sensory nouns, and we speculate that are several possible interpretations. One of the most straightforward hypotheses is that the perception of endurant targets is instantaneous (as it does not have to involve time), while the perception of perdurant targets requires experiencing time that naturally adds to processing latency. It may also be attributed to taking the SNAP perspective (which is instantaneous) or the SPAN perspective (which requires a time course) (Grenon and Smith, 2004). Lastly, it may also be accounted for in terms of the qualia-based lexical types. That is, perception of natural type properties (i.e. formal and constitutive qualia roles in GL) involves simple (i.e. non-conscious) selection of classes. In contrast, perception of artefactual and complex types (i.e. telic

and agentive qualia roles in GL) involves recalling experiential eventive information (i.e. conscious) integration.

In terms of foundations of language, our theory predicts that all languages should have the noun-verb bifurcation at the foundation of their system of grammatical systems or that there must be two basic categories to instantiate the binary contrast motivated by conceptual (in)dependency of time. A well-known linguistic fact that could pose a serious challenge to this claim involves tense-marked nouns.[11] The challenge is that if nouns are indeed endurant and defined without reference to time, how can they be marked by time through tense and aspect?

Nordlinger and Sadler (2004) provided a comprehensive set of data from various languages showing that nouns can be marked by tense/aspect and introduced several important theoretical issues. A debate ensued on how to account for this phenomenon and its associated theoretical implications (e.g. Nordlinger and Sadler, 2004, 2008; Tonhauser, 2007, 2008). The debate focused on issues such as whether nominal tense can be applied to the whole sentence or if it is limited to the local context of the tense-marked noun (e.g. clause/phrase). Bertinetto (2020) proposed a comprehensive account that treats such tense makers as part of the more comprehensive set of nominal semantic features that were originally thought to be verbal features and specifically brings out the temporality of nouns, such as being engaged in a specific event or possessing a specific ability.

Recall that our basic claim of nouns being endurant is based on the fact that each occurrence of the same noun is considered an instantiation of the same entity despite obvious changes in the significant features and or environment of that entity. In contrast, verbs being perdurant and being occurrents have to do with the fact that each occurrence of the same verb (at different spatio-temporal locations) is considered a separate event. Tense-marked nouns do not affect this fundamental dichotomy at all. In fact, tense markers only underline the endurant features of nouns by showing that instances of the same noun retain the same reference regardless of the differences in the explicit information of temporality that they may be carrying. Consider the following example sentence from Huang (2015):

11.	Yī	gōngjīn	ròu	zhǔ shú	hòu	zhǐ	shèng	bù	dào	600	gōngkè.
	one	kilogram	meat	cook-ed	after	only	left	not	arrive	600	gram

'One kilogram of meat only weighs less than 600 grams after being cooked.'

The above example can be viewed in the same spirit as Bertinetto's (2020) account and demonstrates that the nominal classifier system (i.e. gōngjīn 'kilogram' and gōngkè 'gram' used to modify meat) can also be used to mark temporality. The classifier system in Chinese is generally considered to consist of two main subcategories, individual classifiers and measure words (e.g. Ahrens and Huang, 2016). Huang (2015) showed that the individual classifiers mark the endurant feature of nouns, being required to remain identical for the same entity. Measures (hence measure words) are, however, sensitive to spatio-temporal context. Thus, the same noun/entity can be modified by different measurements even though the modified nouns remain "the same," as in retaining the same reference, as shown in the example above. In sum, tense markers of nouns, just like measure words and DE-insertion discussed in Huang (2015, 2016), are simply linguistic devices that allow a language to highlight the variations of the same endurant entity and which would not be interpretable if the modified noun is not endurant.

Conclusion
This paper is the first to establish the cognitive motivation of noun vs. verb bifurcation without presupposing any prior knowledge of grammatical categories. In particular, inspired both by Aristotle's definition that nouns make no reference to time and the more

recent Aristotelian primary ontological bifurcation of endurant vs. perdurant, we propose that the manipulation of the ontological perspectives to obtain time (in)dependent conceptualisation is the foundation of human cognition and of grammatical categories.

To verify this hypothesis, this study focuses on the sensory nouns of the five sensory modalities and the analysis was carried out according to the three lexical types and their associated qualia structures as elaborated in the GL theory. Our findings demonstrate that the time-independent/endurant or time-dependent/perdurant concepts are encoded differently in sensory nouns. Such disparity further differentiates cognitive properties of sensory modalities in the light of embodied cognition. Tactile entities, nearly always endurant, support the intuition of why touch is considered the most embodied (i.e. closely related to bodily contact and involvement), or the most concrete sense, among the sensory modalities (e.g. Zhao et al., 2019). Hearing is the least embodied or the most abstract sense, and its sensory properties are dominated by perdurant objects. Although vision is also less embodied, the versatility and dominance of the visual sense render visual objects to evenly encompass endurant and perdurant properties. Because of the sparseness of the data found for the gustatory and olfactory senses, we are not able to propose a more explicit account of the two senses. This, in fact, echoes the lower accessibility and less frequent embodied encoding of olfactory experiences to some extent (Shen, 1997; Shen and Aisenman, 2008). However, the language-specific situation is worth noting because sensory modalities may exhibit different codability patterns in different languages. For example, some languages encode olfactory experiences much more frequently than other senses (Majid et al., 2018), and olfaction plays a critical role in everyday communication in these communities and languages (e.g. Levinson and Majid, 2014; Majid and Burenhult, 2014).

Drawing upon the findings in this study, there are also implications for other related sensory language studies, e.g. linguistic synaesthesia and modality exclusivity norms. For example, words for auditory concepts appear to be the most "exclusive" as found in previous modality exclusivity norms studies (e.g. Chen et al., 2019; Lynott et al., 2020; Zhong et al., 2022), meaning auditory experiences may have little in common with other perceptual experiences; moreover, auditory sensory is considered the most frequent target domain on the scale of the mapping tendency in linguistic synaesthesia (Zhao et al., 2019). We hypothesise that the time-dependent nature of the auditory sense can account for these results to some extent because the auditory sense is the most fluid in terms of its categorical ambiguity among all the sensory modalities (from verbs to nouns). In sum, we propose that the concept of time dependency may drive a possible synergetic account incorporating diverse approaches such as the categorical dependency of meaning mutability, the cognitive basis of parts-of-speech, and the ontological motivation for differences in the linguistic representation of sensory meanings.

Notes
1 From *the Interpretation* by Aristotle, accessed from The Internet Classics Archive http://classics.mit.edu/index.html
2 For example, attribute verbs (verbs that modify nouns in the manner of adjectives) such as *barking* in "a barking dog;" and the verb phrase *wearing a hat* in "The man

wearing a hat is running in the rain." But note that attribute verbs or deverbal adjectives are not the focus of this study.

3 ASP = aspectual markers, including perfective aspects (e.g. *le*) and imperfective aspects (e.g. *zhe*)

4 CL = classifier

5 Note that *bàogào* here is a shortened form of the compound noun *bàogàohuì* 'report meeting/conference,' in which *huì* 'meeting' is the head in this compound structure. As noted in Wang (2013), most event nouns in Chinese are compound nouns.

6 Lexical typing structure was previously named "lexical inheritance structure" in Pustejovsky (1991, p. 419; 1995, p. 61).

7 *Chinese WordNet* is a platform providing an ontological network of semantic meanings of a particular word coupled with their semantic relations, including hypernyms, hyponyms, synonyms, among others. Accessed at http://lope.linguistics. ntu.edu.tw/cwn2/.

8 *Chinese Web 2011 (zhTenTen11)* is an annotated corpus consisting of a total of 1.7 billion web data crawled in 2011. Accessed at https://the.sketchengine.co.uk/auth/ corpora/.

9 Word Sketch function is a summary of a word's grammatical and collocational behaviour in the *Sketch Engine*.

10 Only those performing arts with oral sounds can be selected. For example, GUO Degang is known as a crosstalk comedian, and crosstalk relies on sound producing when it is performed.

11 We would like to thank Professor Mary Dalrymple and an anonymous reviewer for raising this important issue. Responsibilities for any potential errors are ours.

References

Ahrens K (1999) The mutability of noun and verb meaning. In: Yin Y-M, Yang I-L, Chan H-C (Eds.) Chinese language and linguistics V: symposium series of the Institute of Linguistics (Preparatory Office). Academia Sinica, Taipei, pp. 335–371

Ahrens K, Huang C-R (2016) Classifiers. In: Huang C-R, Shi DX (eds) A reference grammar of Chinese. Cambridge University Press, pp. 169–198

Arp R, Smith B, Spear AD (2015) Building ontologies with basic formal ontology. The MIT Press

Baker MC (2003) Lexical categories: verbs, nouns, and adjectives. Cambridge University Press

Bertinetto PM (2020) On nominal tense. Linguist Typol 24(2):311–352. https://doi. org/10.1515/lingty-2020-2033

Chen I-H, Zhao Q, Long Y, Lu Q, Huang C-R (2019) Mandarin Chinese modality exclusivity norms. PLoS ONE 14(2):e0211336. https://doi.org/10.1371/ journal.pone.0211336

Chersoni E, Santus E, Huang C-R, Lenci A (2021) Decoding word embeddings with brain-based semantic features. Comput Linguist 47(3):1–36. https://doi.org/ 10.1162/coli_a_00412

Chierchia G (1986) Topics in the syntax and semantics of infinitives and gerunds. Ph.D. thesis, University of Massachusetts, Ann Arbor, MI

Chomsky N (1970) Remarks on nominalization. In: Jacobs RA, Rosenbaum PS (eds) Readings in English transformational grammar. Ginn, pp. 184–221

Clark A (1997) Being there: putting brain, body, and world together again. MIT Press

Deng D (2021) Definition of event nouns in contemporary Chinese and the related issues [Xiandai Hanyu shijian mingci de jieding ji xiangguan wenti]. Lexicogr Stud [Cishu Yanjiu] 4:80–91

Firth JR (1962) A synopsis of linguistic theory, 1930–1955. In: Firth JR (ed) Studies in linguistic analysis. Basil Blackwell, pp. 1–32

Fu J (1994) On deriving Chinese derived nominals: evidence for V-to-N raising. Ph.D. thesis, University of Massachusetts, Amherst

Gangemi A, Guarino N, Masolo C, Oltramari A, Schneider L (2002) Sweetening ontologies with DOLCE. Springer, Berlin, Heidelberg, pp. 166–181

Gentner D (1978) On relational meaning: the acquisition of verb meaning. Child Dev 49(4):988–998. https://doi.org/10.2307/1128738

Gentner D (1982) Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In: Kuczaj SA II (ed) Language development, vol 2: Language, thought and culture. Lawrence Erlbaum Assoc Inc

Gentner D, France IM (1988) The verb mutability effect: studies of the combinatorial semantics of nouns and verbs. In: Adriaens G, Small SL, Cottrell GW, Tanenhaus MK (eds) Lexical ambiguity resolution: perspectives from psycholinguistics, neuropsychology, and artificial intelligence. Kaufmann, pp. 343–382

Givón T (2001) Syntax: an introduction, vol I. John Benjamins Publishing Company

Grenon P, Smith B (2004) SNAP and SPAN: towards dynamic spatial ontology. Spat Cogn Comput 4(1):69–104. https://doi.org/10.1207/s15427633scc0401_5

Grimshaw JB (1990) Argument structure. MIT Press

Haegeman LMV (1994) Introduction to government and binding theory, 2nd edn. Blackwell

Han L (2010) Shijian mingci yanjiu zonglun [Overview on the research of event nouns]. Zaozhuang Xueyuan Xuebao [J Zaozhuang Univ] 27(1):86–92. https://doi.org/10.3969/j.issn.1004-7077.2010.01.019

Hsieh S-K, Hong J-F, Huang C-R (2022) The extreme poverty of affixation in Chinese: rarely derivational and hardly affixational. In: Huang C-R, Lin Y-H, Chen I-H, Hsu Y-Y (eds) The Cambridge handbook of Chinese linguistics. Cambridge University Press

Huang C-R (2015) Notes on Chinese grammar and ontology: the endurant/perdurant dichotomy and Mandarin D-M compounds. Lingua Sin 1(1):1–22. https://doi.org/10.1186/s40655-015-0004-6

Huang C-R (2016) Endurant vs. perdurant: ontological motivation for language variations. In: Park JC, Chung J-W (eds) Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation (PACLIC-30), Seoul, Korea. Association for Computational Linguistics, pp. 15–26

Huang C-R, Ahrens K (2003) Individuals, kinds and events: classifier coercion of nouns. Language Sci 25(4):353–373. https://doi.org/10.1016/S0388-0001(02)00021-9

Huang C-R, Calzolari N, Gangemi A, Lenci A, Oltramari A, Prévot L (2010a) Ontology and the lexicon: a natural language processing perspective. Cambridge University Press

Huang C-R, Hsieh S-K, Hong J-F, Chen Y-Z, Su I-L, Chen Y-X, Huang S-W (2010b) Chinese Wordnet: design, implementation, and application of an infrastructure for cross-lingual knowledge processing. J Chin Inf Process 24(2):14–23

Kilgarriff A, Baisa V, Bušta J, Jakubíček M, Kovář V, Michelfeit J, Rychlý P, Suchomel V (2014) The sketch engine: ten years on. Lexicography 1(1):7–36. https://doi.org/10.1007/s40607-014-0009-9

Kwong OY, Tsou BK (2003a) Categorial fluidity in Chinese and its implications for part-of-speech tagging. Research Note Sessions of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL-03), Budapest, Hungary

Kwong OY, Tsou BK (2003b) A synchronous corpus-based study of verb–noun fluidity in Chinese. In: Ji D, Kim T (eds) Proceedings of the 17th Pacific Asia Conference on Language, Information and Computation (PACLIC-17), Sentosa, Singapore. Colips Publications, pp. 194-203

Langacker RW (1987) Foundations of cognitive grammar. Stanford University Press

Levinson SC, Majid A (2014) Differential ineffability and the senses. Mind Language 29(4):407–427. https://doi.org/10.1111/mila.12057

Lynott D, Connell L (2013) Modality exclusivity norms for 400 nouns: the relationship between perceptual experience and surface word form. Behav Res Methods 45(2):516–526. https://doi.org/10.3758/s13428-012-0267-0

Lynott D, Connell L, Brysbaert M, Brand J, Carney J (2020) The Lancaster Sensorimotor Norms: multidimensional measures of perceptual and action strength for 40,000 English words. Behav Res Methods 52(3):1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Majid A, Burenhult N (2014) Odors are expressible in language, as long as you speak the right language. Cognition 130(2):266–270. https://doi.org/10.1016/j.cognition.2013.11.004

Majid A, Roberts SG, Cilissen L, Emmorey K, Nicodemus B, O'Grady L, Vos CD, Dingemanse M, Brown P, Hill CEB, Sicoli MA, Levinson SC (2018) Differential coding of perception in the world's languages. Proc Natl Acad Sci USA115(45):11369–11376. https://doi.org/10.1073/pnas.1720419115

Nordlinger R, Sadler L (2004) Nominal tense in crosslinguistic perspective. Language 80(4):776–806

Nordlinger R, Sadler L (2008) When Is a temporal marker not a tense? Reply to Tonhauser 2007. Language 84(2):325–331

Pustejovsky J (1991) The generative lexicon. Comput Linguist 17(4):409–441

Pustejovsky J (1995) The generative lexicon. MIT Press

Pustejovsky J (2001) Type construction and the logic of concepts. In: Bouillon P, Busa F (eds) The language of word meaning. Cambridge University Press, pp. 91–123

Pustejovsky J (2013) Type theory and lexical decomposition. In: Pustejovsky J, Bouillon P, Isahara H, Kanzaki K, Lee C (eds) Advances in generative lexicon theory. Springer, Dordrecht, pp. 9–38

Pustejovsky J, Havasi C, Littman J, Rumshisky A, Verhagen M (2006) Towards a generative lexical resource: the Brandeis Semantic Ontology. In: Calzolari N, Choukri K, Gangemi A, Maegaard B, Mariani J, Odijk J, Tapias D (eds) Proceedings of the fifth international conference on Language Resources and Evaluation (LREC'06). European Language Resources Association, pp. 1702–1705

Pustejovsky J, Jezek E (2008) Semantic coercion in language: beyond distributional analysis. Ital J Linguist 20(1):175–208

Pustejovsky J, Jezek E (2014) Introducing qualia structure. In: Pustejovsky J, Jezek E (eds) A guided to generative lexicon theory. Oxford University Press

Redington M, Chater N, Huang C-R, Chang L-P, Finch S, Chen K-J (1995) The universality of simple distributional methods: identifying syntactic categories in Mandarin Chinese. In: Proceedings of the international conference on cognitive science and natural language processing. Dublin City University

Redington M, Crater N, Finch S (1998) Distributional information: a powerful cue for acquiring syntactic categories. Cogn Sci 22(4):425–469. https://doi.org/10.1016/S0364-0213(99)80046-9

Sanchez G, Hartmann T, Fuscà M, Demarchi G, Weisz N (2020) Decoding across sensory modalities reveals common supramodal signatures of conscious perception. Proc Natl Acad Sci USA 117(13):7437–7446. https://doi.org/10.1073/pnas.1912584117

Shao J, Liu Y (2001) Lun mingci de dongtaixing ji qi jiance fangfa [Discussion on dynamic property of nouns and the identification method]. Hanyu Xuexi [Chin Lang Learn] 6:1–6. https://doi.org/10.3969/j.issn.1003-7365.2001.06.001

Shen Y (1997) Cognitive constraints on poetic figures. Cogn Linguist (includes Cogn Linguist Bibliogr) 8(1):33. https://doi.org/10.1515/cogl.1997.8.1.33

Shen Y, Aisenman R (2008) 'Heard melodies are sweet, but those unheard are sweeter': synaesthetic metaphors and cognition. Language Lit 17(2):107–121. https://doi.org/10.1177/0963947007088222

Simons P, Melia J (2000) Continuants and occurrents. Proc Aristot Soc Suppl Vol 74:59–92

Song ZY, Huang C-R (2018) Shengcheng ciku lilun yu Hanyu yanjiu [The generative lexicon: studies on the Chinese language]. The Commercial Press

Strik Lievers F, Winter B (2018) Sensory language across lexical categories. Lingua 204:45–61. https://doi.org/10.1016/j.lingua.2017.11.002

Tonhauser J (2007) Nominal tense? The meaning of Guaraní nominal temporal markers. Language 83(4):831–869

Tonhauser J (2008) Defining crosslinguistic categories: the case of nominal tense (reply to Nordlinger and Sadler). Language 84(2):332–342

Wang S (2013) Semantics of event nouns. Ph.D. thesis, The Hong Kong Polytechnic University, Hong Kong

Wilson M (2002) Six views of embodied cognition. Psychon Bull Rev 9(4):625–636. https://doi.org/10.3758/BF03196322

Zhao Q, Huang C-R, Ahrens K (2019) Directionality of linguistic synesthesia in Mandarin: a corpus-based study. Lingua 232 https://doi.org/10.1016/j.lingua.2019.102744

Zhong Y, Wan M, Ahrens K, Huang C-R (2022) Sensorimotor norms for Chinese nouns and their relationship with orthographic and semantic variables. Language Cogn Neurosci. https://doi.org/10.1080/23273798.2022.2035416

Acknowledgements

Competing interests

The authors declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Correspondence and requests for materials should be addressed to Chu-Ren Huang.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.