

汉字所表达的知识系统： 意符为基本概念导向的事件结构^{*}

黄居仁 香港理工大学 洪嘉麒 台湾师范大学
陈圣怡 台湾中研院 周亚民 台湾台北大学

提要 《说文解字》以意符作为汉字的构字要件并发展成部首分类基础。本文以《说文》意符的意义作为造字时所表达的基本概念,分析原意符与其所衍生的汉字的意义关系,建构一个完整的知识体系;并在亚里士多德的经验架构及后续 Pustejovsky 的“衍生词汇理论”的基础上,验证此假设。本研究厘清《说文》“+”、五官类意符(目、耳、口、鼻、舌)与其所有从属字间的概念关系,建构“+”、“五官类”基本概念带领的个别意符知识系统。本文研究显示以“+”为意符的知识系统,所从属词汇与其性质相关;以“五官类”为意符的知识系统皆具共通性,并与五官类单字词的现代词义系统相符。

关键词 意符 知识本体 汉字 词汇语义

1. 汉字系统与知识的关联性

1.1 语言的共有知识: 知识本体与其应用

语言是人类用来传达知识、讯息,与彼此沟通概念的最佳且最主要的工具。在运用语言作为传达、沟通知识的媒介,人们的知识将在共同的认知前提下,达到讯息系统化,也唯有如此,人们的知识才得以跨越时间、空间的交替而传承、保存。黄居仁(2009)提及使用同一语言的人,用该语言作为表达、传递,并共享知识的架构。因此,该语言背后所依赖的概念与知识架构,必须是所有会讲该语言的人都共有的,也就是现今许多学者所重视的知识语言,也就是知识本体(ontology)。而所谓的“知识本体”无法表达细部知识,只能表达概念结构。知识本体的架构越抽象、层次越高,其所涵盖的知识范围越广;但相对的,它离实用的知识层次越远,能表达的实际知识越少。语言涵盖了所有人类所能表达的知识,因此,任何语言都有其内涵的知识本体,所有讲这个语言的人,都不自觉地在用这个隐含的约定俗成系统(黄居仁等2010a)。

从汉字意符的概念与特质出发,厘清汉字意符与其所从属部件之间复杂的

* 本文为黄居仁中研院深根计划奖助之研究成果。陈圣怡进行本研究时任职于中研院语言学研究所,目前已经不在该单位服务。

关系，以知识本体架构的观点，来探讨汉字内涵的知识系统。在本研究中，我们试图藉由这样的知识本体框架来传达汉字所要表达其内涵的知识系统。

1.2 由汉字意符出发的知识体系架构

语言的主要功能，就是传递表达讯息。而当这些讯息经过系统化成为知识，语言也自然成为承载知识的主要媒介。从知识本体的研究中，我们看出，承载知识的表达系统，本身必然也有严谨的知识体系（黄居仁等 2010a）。换句话说，语言之所以能作为有效的沟通工具，是因为使用同一语言的人，必须同时接受该语言内涵的知识系统。在这个前提下，知识系统语言间的接口，变得十分重要，这也是黄居仁等（Huang, et al. 2010a）一书所代表的新研究方向，称为“本体-词汇接口”（OntoLex interface）。在这个研究方向中，核心的议题是如何发现并运用语言中约定俗成的知识系统。

汉字书写系统是非常值得深入研究的数据，汉字的构字特性乃是以意符所表达的基本概念为中心，汉字书写形式和字义大半皆可由意符概念而衍生。因此，意符表达的知识系统是汉字书写系统的核心。周亚民、黄居仁（2005）提出，汉字与其他书写系统相比，最大的差异是汉字有很好的表意能力，字符（glyphs）与概念的连结强，有时候不一定要知道发音，也可以从字符知道大概的意义，本身就构成一个知识结构，但拼音文字和音节文字与概念的连结，是透过发音对映到意义，字形与概念的连结弱，汉字的表意特性来自于字符结构中的表意符号，而且这些意符的表意能力，历经数千年后仍保存得相当好。

此外，黄居仁等（Huang, et al. 2008a, 2008b）以及陈圣恰等（Chen, et al. 2008），透过对于汉字意符所要表达的基本概念与其衍生意义，他们皆对汉字书写系统的架构有深入的探讨，使我们更能理解汉字书写系统的完整性与科学性，并能理解汉字意符与其从属字之间的关系连结性。

汉字表达的意义都与所使用的意符有概念上的联系，而大部分汉字的意义，都是意符所表达概念的延伸，或是词义的扩展或缩小，因此，如果掌握意符的知识结构，等于掌握了大部分汉字意义，我们希望建立的是整体的知识架构，而不是各别意符的片断。我们尝试以《说文解字》的五百四十部首作为汉字的基本意符，分析部首作为意符时所表达的概念，并运用 IEEE “建议上层共享知识本体”（Suggested Upper Merged Ontology, SUMO）表达并呈现意符的知识系统。

2. 共享上层架构的知识表达：SUMO

为了追求语言知识架构的丰富性，我们采用 SUMO。由于不同知识系统对各种概念与名词，可能由不同的定义与定位，无法完全兼容。因此知识系统之间的沟通，需要一个共有的上层知识系统，IEEE 于是成立一个工作小组——标

准上层知识本体工作小组，并建立了一个 SUMO (Niles and Pease 2001) 。SUMO 采用“标准上层本体知识交换格式”(Standard Upper Ontology Knowledge Interchange Format , SUO-KIF) 来描述概念间的关系。

上层知识本体是将一般性、后设性(meta) 、摘要性以及哲学类词汇的概念指出，所以特殊领域词汇的概念可由其中的概念所涵盖，但特殊领域概念的知识本体则期许由各领域自行制订(Niles and Pease 2001 , 2003) 。目前 SUMO 已经和英语词汇网络 WordNet1.6 及 2.0 版本作连结，使得任何领域的知识都可以借助词汇的连结，建立正确的知识本体位置。完整的 SUMO 数据库，可以在其官方网站 <http://www.ontologyportal.org> 上检索或下载。

SUMO 的中文化，最初是依据 SUMO 2002 年版的数据进行系统接口及概念节点的中文化，由中研院语言所中文词网小组完成，内容主要参考 Wordnet1.6 的英中对译。之后，于 2004 年进行 termformat 及 format 的中文翻译修正。中央研究院中英双语知识本体词网 (The Academia Sinica Bilingual Ontological WordNet , Sinica BOW <http://bow.sinica.edu.tw> , Huang , et al. 2004 , 2010b) 已进一步将中文词汇、WordNet 和 SUMO 之间建立连结。因此，我们可在 Sinica BOW 上直接以中文词汇查询相对的 SUMO 知识本体概念。

3. 汉字知识本体

汉字书写系统的特殊性，在于其以“意符”为中心的构字方式。因此，我们认为汉字的意符代表了汉字知识系统中的基本概念，整个汉字意符的系统表现出来的即是一个知识分类的体系。早在公元 121 年，许慎的《说文解字》就将汉字依照意符分为 540 个类别，一般称为 540 个部首。虽然，当时的许慎并无知识本体的概念，但他所建立的 540 部首，确实在一定程度上反映了造字时所表达的基本概念。

汉字意符的知识结构是什么，至今仍然没有找出完整的结构，最主要的问题是缺乏一个可以作为不同语言知识结构比较的共同知识体系。但是近年来上层知识本体 (upper ontology) 的建立，给我们提供了研究汉字意符知识结构的可能。周亚民的博士论文(2005) 将《说文解字》540 个部首的本义与 IEEE SUMO 对应，建立了表达汉字意符知识概念的基础结构。此外，周亚民、黄居仁 (2013) 对于汉字知识本体的架构、建构步骤以及汉字意符知识结构的建立，皆有详尽的探讨。我们即以其研究成果作为建构“汉字知识本体”的基础，探讨汉字意符知识系统。

4. 意符是汉字的构字要件

《说文解字》(许慎 2000 [121]) 更以意符作为汉字部首分类的基础。近年计算词汇语意学的研究中，已尝试将汉字意符的语意分类系统与词网或知识

本体结合(如 Wong and Pala 2002; Hsieh 2006 等)。这些相关研究中,又以周亚民(2005)最为完整,他将《说文解字》540 个部首的本义与 SUMO (Pease and Niles 2001) 对应,建立了表达汉字意符知识概念的基础结构。而后并发展成有描述汉字历史演变的能力系统(周亚民、黄居仁 2006),在这系统的基础上,周亚民、黄居仁(2005)进一步观察到汉字意符具有与衍生词汇(Pustejovsky 1995)相似的衍生能力。

本文遵循意符代表基本概念这个主张,并希望透过进一步的分析,将所有从属字与这个基本概念间的关系厘清,以建构成一个由该基本概念带领的知识系统。每一个意符知识系统可看成一个由该基本概念领头的领域知识本体。

单一意符所代表的基本概念要如何透过概念衍生而形成一个完整的知识体系?其知识衍生的架构为何呢?希腊哲学家亚里士多德曾提出知识的经验架构(qualia),近来为 Pustejovsky (1995) 的“衍生词汇理论”所引用。此理论指出人类知识的衍生,主要是由物质(formal)、组成(constitutive)、功用(telic)与产生(agentive)等四个面向而来。本研究根据实际分析的结果,进一步将汉字意符的衍生面向,归纳为物质、组成、功用、事件(participating)、参与者(participant)、描述状态(descriptive)与产生等七大类。

本研究以同属于 SUMO“躯体部件”类别下的意符“+”、“五官类:目、耳、口、鼻、舌”等为研究对象,尝试建构其概念衍生结构,藉以观察类别性质相似的意符,其概念衍生方向的特色与异同。此外,更进一步将“+”、“五官类”意符的概念衍生结构与这些单字词在中文词汇网络(Chinese WordNet)中词的现代词义对照,以观察汉字意符的知识结构与其现代词义之间的相关性。

5. 意符知识系统介绍

本研究的主要工作接口是中研院词网小组开发的意符知识系统。意符知识系统是一套表达汉字基本概念与其从属字之间关系的知识本体架构。本系统延伸周亚民(2005)将《说文解字》540 个部首本义与 IEEE SUMO 对应的基础结构,根据进一步的概念衍生分析结果,表达每个部首的知识架构体系。

5.1 部首查询

“意符知识系统”提供两种部首查询方式:

(1) 依 SUMO 概念分类查询。选取特定 SUMO 概念,则包含此概念的下层 SUMO 概念也会出现供使用者查询。例如选“动物”类,会出现下层的“哺乳类”、“脊椎动物”、“动物”等概念部首以供查询。

(2) 依部首字形查询。输入字形,可直接查询该部首数据。

5.2 基本概念

依照《说文》对部首的释义,及本研究对部首下从属字的分析结果,可归纳

出部首意符的基本概念及其下层的衍生概念。例如“羊”的基本概念为“有啼哺乳动物”。“目”的基本概念为“身体器官”；下层衍生概念为“视觉”。

5.3 汉字意符衍生概念分类

根据《说文》释义，本架构将从属字与部首基本概念的相关依照分类表达出来。本研究根据 Pustejovsky (1995) 衍生词汇理论的“环境-经验”(qualia) 为基础，再加上对部首从属字的实际分析结果，将汉字意符的衍生面向，区分为七大类。

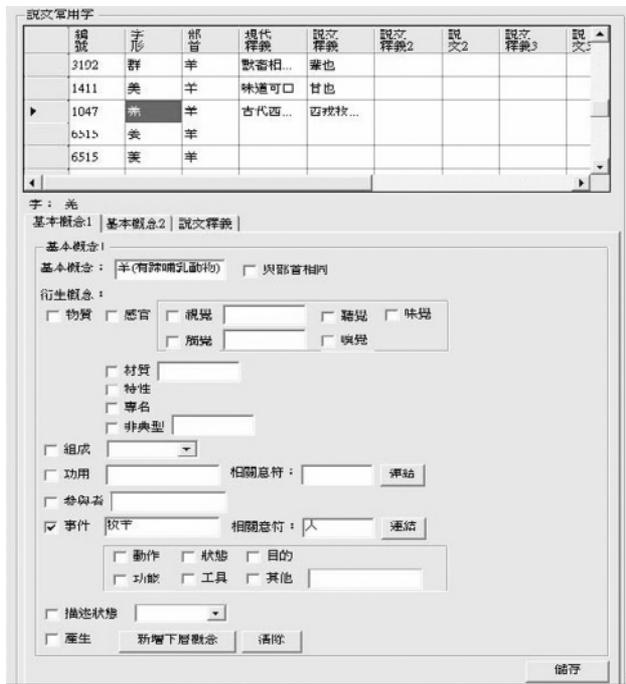


图1 汉字意符衍生概念例示 “美”的事件是“牧羊”

(1) 物质。物质下分为“感官、材质、特性、专名、非典型”五类。“感官”下又分为“视觉、听觉、嗅觉、味觉、触觉”五个类别。凡释义对从属字的描述是来自感官经验的，皆归属于“感官”类。基本概念作为从属字材质的，归属于“材质”类。描述从属字特性的，归属于“特性”类，例如“骏，马之良材者。”释义为名称的归属于“专名”类。例如“驩，马名。”与基本概念没有办法直接产生连结的，归于“非典型”类。例如“驴，驴兽。似马，长牙。”驴在马部底下，却不属于马的一种，于是归属“非典型”类。

(2) 组成。下有“部位”、“分子”、“整体”三个类别。车部的“辑”释义为“辑，车輿也。”车輿是车的一个部位，归属于“部位”类。马部的“羴”字释义为“羴，众马也。”以数量、整体观点来看，众马是马的整体，也就是说，

马是众马的一部分，归于“分子”类。

(3) 功用。车部的“输”字释义为“输，委输也”。运输是车的功用，归属于“功用”类。

(4) 参与者。可表现出与释义描述内容相关的参与者。例如马部的“驱”释义为“驱马也。”在驱马的事件中，“人”是参与者。

(5) 事件。依照事件形态，又可分为“动作、状态、目的、功能、工具、其他”六个细类。如图1所示，“羊”是“羌”所描述的“牧羊”事件的对象之一。

(6) 描述状态。依照状态形态又细分为“动态”与“静态”两类。例如马部的“驯”释义为“马顺也。”归类为描述状态的“静态”；马部的“馱”释义为“马摇头也。”归类为描述状态的“动态”。

(7) 产生。从属字与意符的关系来自生产关系的，归于“产生”类。例如：羊部的“𦍋”释义为“五月生羔也。”

5.4 相关意符连结

“功用”与“事件”的类别下，有“相关意符”的字段，可表示出与衍生概念相关的意符，并将二者作连结。例如羊部的“羌”释义为“西戎牧羊人也。”牧羊表现的是一个事件，“人”作为牧羊事件的参与者，在“羌”的构字中，也可看出“人”是其相关意符。系统将“羌”与其相关意符“人”作连结，可供查询时交互参考之用。



图2 汉字“羌”的相关意符连结是“人”

6. 意符的概念衍生架构与特色

6.1 “++”意符

6.1.1 “++”意符的概念衍生架构

周亚民和黄居仁(Chou and Huang 2010)在汉字意符的研究中提到,意符“++”的衍生字和衍生概念有植物的专名、描述植物的部位、描述植物的属性和外观,以及描述植物的功用,如图3所示,这个知识体系符合由Pustejovsky提出的“衍生词汇理论”,这个理论说明了词汇衍生的环境-经验结构,此结构可分为四个面向:物质、组成、功用、产生,由意符“++”的衍生概念来看,正好符合物质(植物)、组成(植物的部分)、功用(植物的利用),而植物非人造物,所以没有“产生”此面向的衍生概念.由研究结果来看,Chou和Huang(2010)所建立的意符概念结构可以证明意符具有与衍生词汇相似的衍生能力。

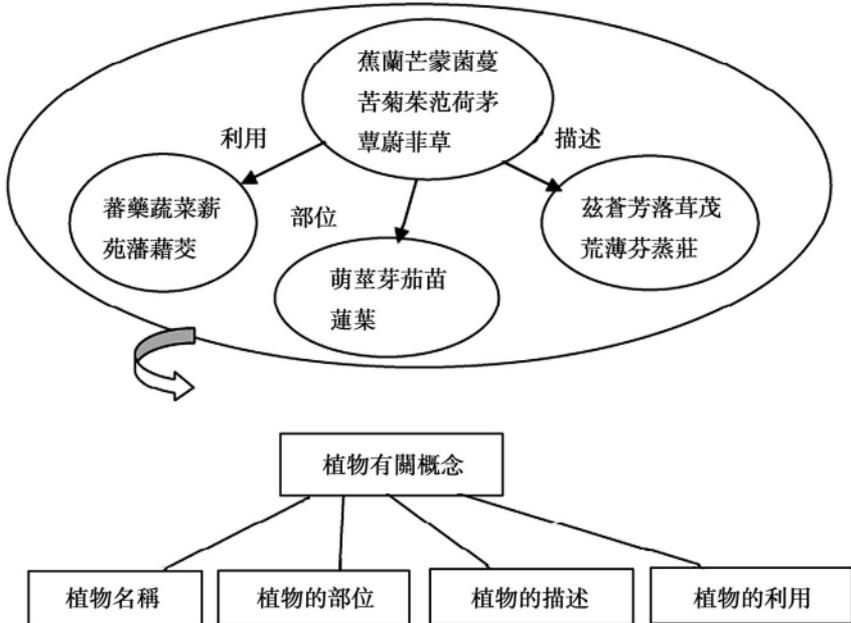


图3 意符“++”的衍生概念示意图

6.1.2 “++”意符的的衍生与现代词义

以“++”为意符的分析,我们发现透过“衍生词汇理论”,与植物有关的概念,可分为:植物名称、植物的部位、植物的描述、植物的利用。例如,在“植物的部分”,所衍生出来的词汇就是在植物不同位置出现的物质,有“茎、芽、苗、叶……”等字,因此,我们在教学应用上,就可以教导学习者,植物的主干,下部与根连结,上部有花、果、叶,就是“茎”;植物的初生苗,就是“芽”;植物的一部分,生在枝干上,专营吸收、蒸发等作用,就是“叶”。“植物的描述”,是描述与植物相关的状态,进而衍生到描述一般事物或一般状态,

“茂”指的是植物繁盛、旺盛，衍生为繁多，后引申为丰富优美的事物或状态；“芳、芬”本指植物的气味，衍生为香气，再引申为指美好的事物。“卅”意符的汉字教学应用，先以“卅”意符为出发，透过经验结构的衍生，达到了解相关词汇的词义。除了具有科学性、系统性的教学应用，也提供学习者一套据有关联性的汉字学习系统，以了解、相关性来理解汉字，取代以死背的方式来学习汉字。

以“药”为例，是一个常用，又不是很容易学的字，可以表示“药物”、也可以表示“火药”，两者的原始义皆表示其材料取自与“卅”相关的物质，并利用这些与“卅”相关的物质加工、改制而成，因此，根据图3所示，“药”的归类，就是在以意符“卅”的衍生字与衍生概念的“利用”这一类。

6.2 “五官类”意符

何谓“五官类”？“五官”这个词被使用得很广，在佛教、中医、面相学等不同领域中，对“五官”亦各有不尽相同的定义。进一步查阅辞典，商务印书馆的《现代汉语词典》(2005年第5版)对“五官”的解释则是“指耳、目、口、鼻、舌”。“耳、目、口、鼻、舌”正是脸上的五个主要器官，可合称为“五官”。耳主听觉，目主视觉，鼻主嗅觉，舌主味觉，口则与饮食和言语有关。这五个器官是人类认识世界、与世界互动、沟通的重要媒介。本研究即以《说文》意符中的“耳、目、口、鼻、舌”五个意符为研究对象，分析“五官类”意符在产生从属字时的概念衍生方向，以建构“五官类”意符知识系统的架构，并从而探讨在语言中呈现的“认知”概念体系。

6.2.1 “五官类”意符的概念衍生架构

以下逐一分析“耳、目、口、鼻、舌”等意符与其从属字间的关系。并透过概念图的形式，建构并表达其知识系统的架构与特色。

6.2.1.1 “目”部

“目”部经过分析后，其从属字的概念衍生面向涵盖了物质、组成、描述状态、事件、功用等五大类，其中又以由功用衍生出的下层概念“视觉活动”的比例最高，其概念衍生图见下页图4。

(1) 属于“物质”类的从属字，大多是对眼睛外观的描述。例如“睥，大目也”，描述眼睛的大小；“盼，白黑分也”，则是描述眼睛黑白分明的样貌。我们分析时将它们归在“视觉”下的“形状”。

(2) “组成”类的从属字，为眼睛可被分析出的相关组成部位。例如：“眦，目匡也”；“睫，目旁毛也”；“睑，目上下睑也”。

(3) “描述状态”类，主要是描述眼睛的静态或动态状态。例如“眊，目少精也”(眼睛昏浊，看不清楚)，“睦，目顺也”(目光祥和)，属于眼睛的静

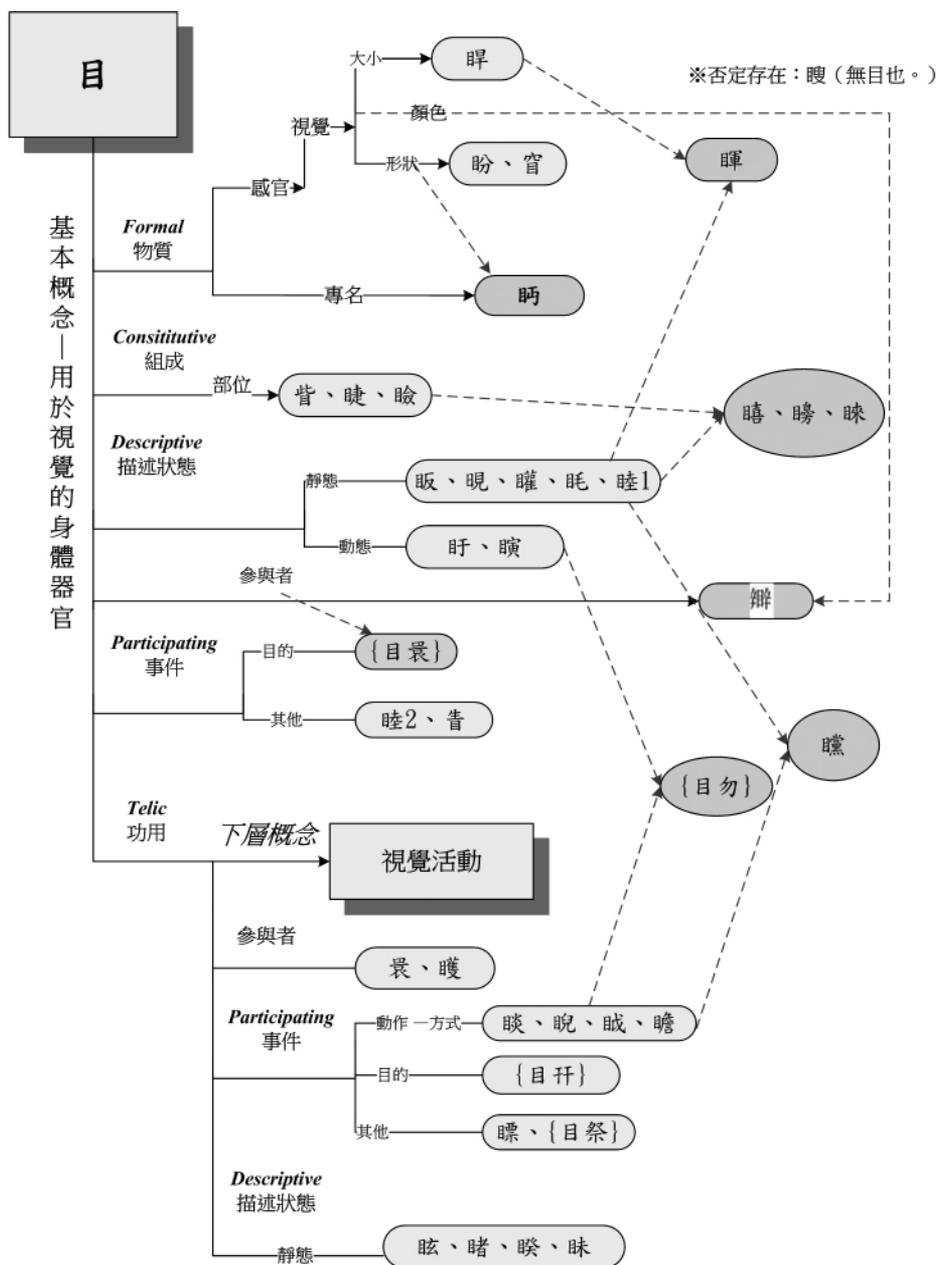


图4 汉字“目”的概念衍生架构示意图

态描述。而“眇，一曰张目也”，“眈，开阖目数摇也”，皆描述了眼睛的动态。

(4) 属于“参与者”的从属字，描述了事件的参与者。例如“辨，小儿白眼也”，参与者为小儿。

(5) “事件”类，皆是与眼睛相关的事件。例如“瞶，儿初生蔽目者”，“瞶”字的概念衍生方向同时来自“参与者”、“事件”两类。参与者为“儿”；事

件为“初生蔽目”，即刚出生时，眼睛遮蔽眼睑。

(6) 属于“功用”类的从属字共有 49 个，在《说文》目部中比例最高，约占 41%。其从属字所描述的对象是“视觉活动”而非“眼睛”本身。因此，我们特别将这一类独立成目部的“下层概念”，以与“目”作为躯体部件的基本概念区隔。由于眼睛首重于“功能”的特性，其下层概念也发展出参与者、事件与描述状态等三大类的概念衍生方向。例如“眗，邪视也”，描述视觉活动的动作方式“昧，目不明也”，则描述了视觉活动的结果是不清晰的。

6.2.1.2 “耳”部

“耳”部的分析结果，其从属字的概念衍生面向涵盖了物质、描述状态、事件、功用等四大类，且有隐喻的概念衍生方向。在这些概念衍生类别中，又以由“功用”衍生出的“听觉活动”所占比例最高，其概念衍生图如图 5。

(1) “耳”部中属于“物质”类的从属字很少，仅有“专名”的用法。例如：“聾，吴楚之外，凡无耳者曰聾也”。本研究将属于方言用法的字归于“专名”类。

(2) “描述状态”类，主要是描述耳朵外观的状态。例如“聵，耳大垂也”；“聾，耳箸颊也”。

(3) 属于“参与者”的从属字，描述了事件的参与者。例如“聾，小儿白眼也”，参与者为小儿。

(4) “事件”类的从属字，又分属于“目的”与“其他”两小类。例如“聵，墮耳也”，耳朵为目标物，因此归于“目的”。另外，“聵，连也。从耳，耳连于颊。从丝，丝连不绝也”。此乃由耳朵与脸颊相连的特性，而进一步引申出表“相连”的抽象词义。凡引申义皆归于“事件”下的“其他”。

(5) 属于“功用”类的从属字，在《说文》耳部中占了一半以上，共有 18 字，为耳部字整体的 55%。反映出与耳朵相关的概念中，以听觉功能为主的特性。其从属字所描述的对象是“听觉活动”而非“耳朵”本身。因此，我们将“听觉活动”独立为耳部的“下层概念”，与“耳”作为躯体部件的基本概念区隔。由“听觉活动”的下层概念也发展出物质、事件与描述状态等三大类的概念衍生方向。例如“声，音也”，声音由听觉而来。而，“闻，知声也”（听见）；“聾，无闻也”（听不见）。听见、听不见，皆为听觉活动的结果，我们将之归于描述状态的静态类。最后，“聾（聖），通也”、“聾，察也”等引申义归入事件中的“其他”。

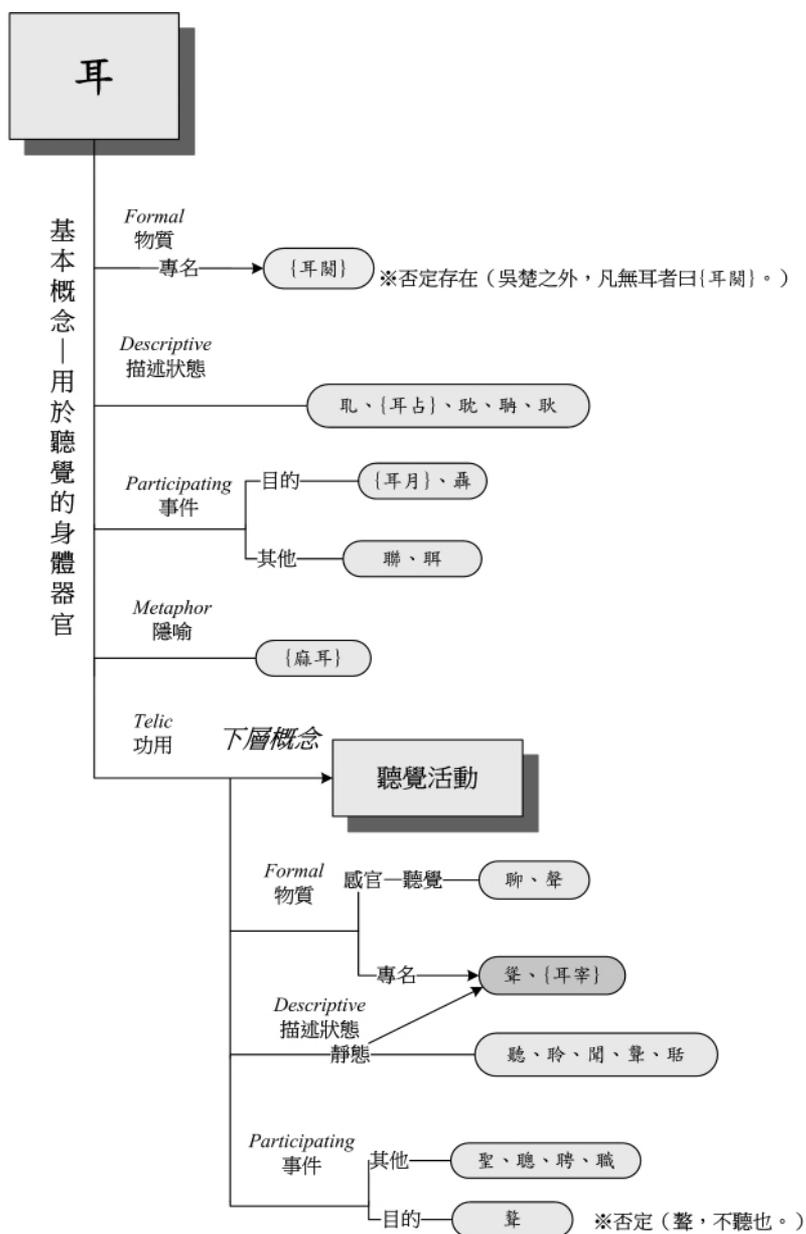


图5 汉字“耳”的概念衍生架构示意图

6.2.1.3 “口”部

“口”部的分析结果，其从属字的概念衍生面向涵盖了物质、组成、描述状态、事件、功用等五大类，且有隐喻的概念衍生方向。口与目、耳部相同，以“功用”衍生出的从属字所占比例最高，其概念衍生图见下页图6。

(1) 由图6可知，口部的概念衍生结构为：以作为身体器官的“口”为基本概念；另外由“口”的两个主要功能——言语和饮食，分别衍生出两个下层概

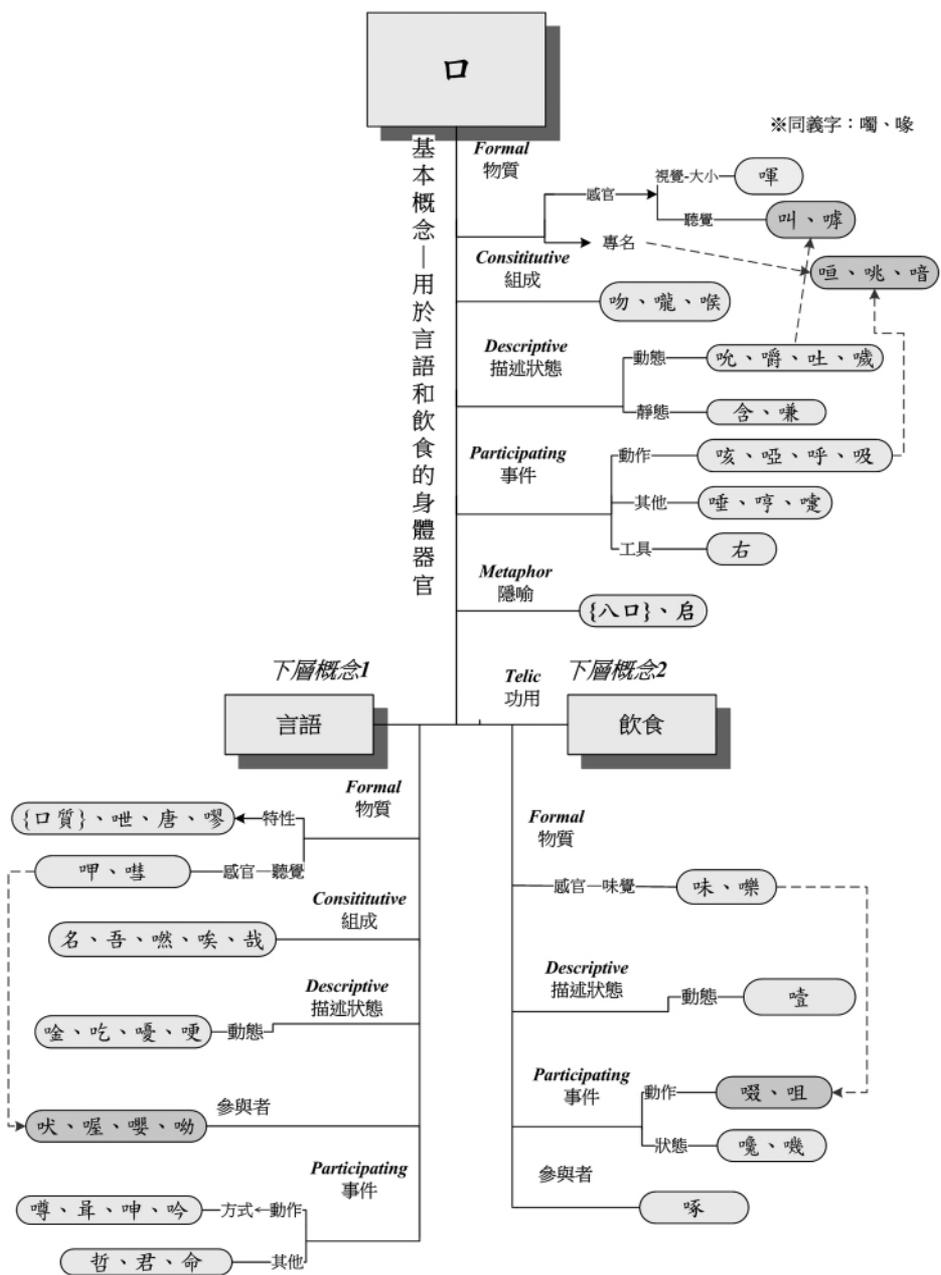


图6 汉字“口”的概念衍生架构示意图

念。且属于下层概念“言语”功能的字有61个，而属“饮食”功能类的从属字有20个，两者加起来共占口部字的47%，将近一半。

(2) “口”部基本概念下的“物质”类，描述了口的外观大小，例如“嘩，大口也”。另外，“叫，噤也”，“嚙，号也”，都与口部发声有关，则归类于物

质下的“听觉”。

(3) “组成”类的从属字，为“口”可被分析出的相关组成部位。例如：“吻，口边也”(嘴唇；“咙，喉也”；“喉，咽也”。

(4) “描述状态”类，主要是描述嘴巴的静态或动态状态。例如“嚼，啮也”，“吐，寫也”(使东西从口里出来)，描述了口部动作的状态。而“含，嗛也”，“嗛，口有所衔”，则为口中含着东西的静态描述。

(5) 属于“事件”类的从属字，又可细分为动作、工具与其他三类。例如：“含，嗛也”，“呼，外息也”，“吸，内息也”，皆是与口部动作有关的事件。而“右，助也，从口又”(手口相助也)。口在“手口相助”的事件中，扮演工具的角色。“唾，口液也”，“嚏，悟解气也”，这些都是与口部事件有关的物质，暂归于“其他”。

(6) “合，山间陷泥地”；“启，开也”。无论是“山间凹陷的泥地”或者“开启”的意义，都与作为身体器官的“口”无直接关系。这两个字皆是透过隐喻而产生出缺口或开口的概念。

(7) 作为“口”的“言语功能”的下层概念，又发展出物质、组成、描述状态、参与者、事件等五个衍生方向；而另一个下层概念“饮食功能”，也发展出物质、描述状态、参与者、事件等四个概念衍生方向。

6.2.1.4 “鼻”部

“鼻”部下的从属字很少，仅有4个字“𪔐，以鼻就臭也”，以鼻子接近气味，这应该是描述“以鼻子闻气味”的动作“𪔐，卧息也”，“𪔐，卧息也”，“𪔐、𪔐”皆表示睡觉时以鼻呼吸“𪔐，病寒鼻塞也”，则是生病时鼻塞不通的事件。由前述可知，鼻部从属字的概念衍生方向皆是由事件而来，而且不论是呼吸或是嗅觉的事件，皆与鼻的功能有关。

6.2.1.5 “舌”部

“舌”部的从属字更少，仅有2个字。虽然“舌”的从属字很少，我们无法建构其意符的知识结构。然而，《说文》释义“舌，在口所以言、别味者也”，则清楚地说明了舌的功能，这一点与其他五官类意符相同。

6.2.2 “五官类”意符知识系统的特色

经由以上对“目、耳、口、鼻、舌”等意符的分析，大致可了解每个意符的概念衍生方向。然而，“鼻”、“舌”部从属字的数量和“目、耳、口”落差太大，也由于字数太少而难以建构其意符概念衍生架构。因此，本文以五官类中的“目、耳、口”这三个概念衍生能力较强、概念发展较完整的三个意符知识架构为主要讨论对象；鼻、舌则作为辅助参考。以下即将各意符之间的概念衍生架构互相比较，可归纳出“五官类”意符知识系统的特色。

(1) “五官类”意符的概念衍生架构十分相似，目、耳、口的衍生概念架构类似，皆是以“身体器官”做为基本概念，再依其各自不同的功能，而进一步衍生出视觉、听觉、言语及饮食等下层概念。

(2) “五官类”意符的以“功用”类的从属字比例最高，除了鼻、舌部的从属字太少，较难观察其从属字的概念衍生类别比例之外，耳部的“功用”类的从属字有 18 字，为耳部字整体的 55%。而目部“功用”类的从属字，共有 49 个，在《说文》目部中比例最高，约占 41%。最后，口部的“功用”类从属字又分为：言语及饮食两大类，属于“言语”功能的字有 61 个，而属“饮食”功能的字有 20 个，两者加起来共 81 字，占口部字的 47%。由上述可知，耳、目、口部的“功用”类从属字，皆约占整体的一半，比例最高。

(3) 《说文》释义与概念衍生方向相符。“目，人眼也。象形。重童子也。”许慎对人眼睛的解释，由外观切入，他描述眼睛的外观，像是层层包覆的瞳子。另外，对耳、口的释义分别是“耳，主听者也。象形”；“口，人所以言食也。象形”。比较前述耳、目、口的说文释义，我们发现一个现象，即“耳”、“口”皆针对其“器官功能”释义；而唯独“目”是由描述眼睛的“外观”来释义。在分析其从属的字的衍生概念后，也发现“目”部在针对器官的外观、动作、状态的从属字比例，确实较“耳”部、“口”部高。由此也可看出，人类对眼睛外观的观察较多、认识也较深。

(4) 意符可以表达隐喻的字义衍生。在“耳、口”部的从属字中，我们发现《说文》已有透过“隐喻”衍生而来的概念。例如“耳”部有“靡，乘舆金耳也”，意即天子乘车上的金饰车耳。“口”部则有“启，开也”，从户从口，表示门有“开口”的概念。

6.2.3 “五官类”意符的概念衍生与现代词义

“目、耳、口、鼻、舌”等单字词，在现代中文的使用中，皆拥有一个以上的丰富词义。这几个词的词义的来源与衍生，是否与其在造字之初的基本概念即有密切的关连？本文试图将“五官类”意符的概念衍生结构与这些单字词的现代词义对照，以观察两者间有何关联性。

6.2.3.1 “五官类”意符的现代词义

在中文词汇网络 (Chinese WordNet, CWN, 黄居仁等 2010b) 中，除了“目”之外，“耳、口、鼻、舌”都已做过词义分析。我们将“耳、口、鼻、舌”的分析结果整理如表 1。

由表 1 中的现代词义整理可知，“口”的词义数最多，有 13 个；“耳”、“鼻”各 3 个，“舌”2 个。现代词义数和意符衍生词义数，都以“口”为最多。其共同特点归纳如下：

(1) 皆以“身体器官”与“器官功能”为主要词义。

(2) 皆有隐喻的词义。例如“形状像舌头的物品”、“形状像耳朵的物体”、“内外相通的出入处”等。

(3) 虽然有《说文》未收入的新词义，但意义衍生的方式，都在意符知识系统架构描述的范围内。例如，“口”在现代出现了许多量词的用法，是较特殊的一点。但有意思的是，这些量词的意义衍生，来自“口”所参与的事件，符合意符“口”原有的词义衍生机制。

耳	0100 普通名词。动物的听觉器官。常用接尾词。 0101 整个听觉器官，包括内部构造。 0102 耳朵表面。 0200 普通名词。动物的听觉。 0300 普通名词。接尾词。形状像耳朵的物体。
口	0100 普通名词。动物的口腔。 0101 普通名词。动物的饮食器官。 0102 普通名词。动物的发声器官。 0103 普通名词。脸部上，口腔表面外型的部分。 0200 位置词。容器供内容物进入的地方。 0300 位置词。江河注入湖海的地方。 0400 位置词。内外相通的出入处。 0500 计算人数的单位。 0600 计算棺材数量的单位。 0700 计算牲畜数量的单位，主要用于“猪”。 0800 计算有大开口器物的单位，常用于“缸”、“瓮”、“钟”。 0900 计算刀、剑的单位。 1000 计算从地面开挖，且可取水或存水的地方的单位，常用于“井”。 1100 计算每次口腔动作所含的量，常用于“食物”、“饮料”。 1200 计算牙齿的集合单位。 1300 准量词。计算言谈的集合单位。
鼻	0100 普通名词。动物呼吸和嗅觉的器官，位于脸部的正中央。 0101 通指整个呼吸和嗅觉的器官。 0102 专指鼻子表面外形的部份。 0200 普通名词。动物的嗅觉。 0300 形容像鼻子一样位于物体最前端的。
舌	0100 普通名词。动物口中辨别味道，帮助咀嚼和发音的味觉器官。 0101 整个舌头，包括其内部组织。 0102 舌头表面外型的部份。 0200 普通名词。接尾词。形状像舌头的物品。

表1 中文词汇网络(CWN)词义分析结果

6.2.3.2 “五官类”意符的概念衍生结构与现代词义

比较意符“耳、口、鼻、舌”的概念衍生结构与“耳、口、鼻、舌”的现代词

义，发现以下几个特点。

(1) 两大主要概念：身体器官与功能。意符耳和口的概念衍生架构，皆是以“身体器官”为基本概念，而“器官功能”为下层概念。其从属字就依循着这两个主要概念衍生、开展。我们也发现，耳、口、鼻、舌的现代词义也兼含了身体器官和功能这两大概念。例如耳(0200 动物的听觉。例如：老人的〈耳〉很背，当我大声说要走时她总是过来摸摸我的头。)；鼻(0200 动物的嗅觉。例如：我的〈鼻〉很灵，要找材料也会快点。)；耳、鼻皆独立出表“感官功能”词义。另外，从词义分析的结果也可看出“口”的现代词义除了是身体器官之外，亦兼具饮食和发声功能等两个重要概念。

(2) 现代词义有更丰富的隐喻用法。在意符“耳、口”部的从属字中，我们发现《说文》已有透过“隐喻”衍生而来的概念，例如“𦉳，乘舆金耳也”；“启，开也”。这几个单字词发展至现代，已经具有更多由隐喻衍生出来的词义。例如“口”就有三个“位置词”的用法(容器供内容物进入的地方；江河注入湖海的地方；内外相通的出入处)，以上三个不同的位置皆可称为“口”，这三个不同的词义，皆透过“口”的隐喻衍生而来。

(3) 量词用法的出现。“五官类”意符的从属字，在《说文》阶段尚未出现“量词”的概念。而在现代词义中，我们发现“口”的量词用法十分丰富，既可以作为计算具体的人数或大开口器物的单位，亦可作为计算抽象言谈的集合单位。

7. 结论

本文从汉字意符系统所蕴含的语言知识出发，发现意符与衍生字族，表现的其实是一个完整的知识系统。这个研究，不但可以提供汉字知识系统完整的描述，其实也为人们将知识系统化的过程提出了最直接的证据纪录。未来的研究中，我们希望能通过更多其他类别意符的分析，以观察不同类别意符在知识结构上的特性与共性，以期能更系统化地建立出汉字意符的知识结构，并提供汉字驱动知识工程与语义运算的基础架构。

本研究主要的贡献是建立汉字新的分类架构，突破了传统文字学对汉字的分类限制，建立了意符的知识架构，并发现意符具有与衍生词汇类似的概念衍生能力。未来我们将会建立不同断代的概念结构，以本研究为基础，进一步比较不同时间的概念变化。

以汉字知识本体系统分析汉字意符的概念，根据 Pustejovsky(1995) 提出的“衍生词汇理论”而发展出的“经验结构”概念所得到的结果，从意符的本义、衍生到引申词义，相关联性的汉字系统，由汉字意符的概念取代传统以部件、偏旁、结构为主的概念，是具有科学、理据、系统的汉字表征知识系统(洪嘉馥、黄居仁 2012)。此外，基本原理的知识本体架构有助于建立一个语义的网

络，虽然这个方法之前曾经被批评有太多的人工标注，但是我们的经验结构知识本体架构的研究方法可以帮助我们建立一个正确的语义相关性连结。

未来对于建立汉字知识系统的发展与趋势，亦可在周亚民(2009)研究的基础上，发展出以数字化系统，结合知识本体的讯息与“衍生词汇理论”的经验结构的概念，由计算机数字化的辅助，提供大量且相关的汉字讯息和语言知识。此外，亦可在这个汉字知识系统的架构下，提供不只是当代的字形、字音、字义和构词，也包含了不同时代的字形、字音、字义、异体字的变化。

总的来说，本研究在汉字意符知识结构的基础之上，进一步以同类别的意符进行类别化的比较研究，由此也发现同类意符在概念衍生结构上，确实具有相似的特性。通过知识本体与词汇语意的研究，也验证了“功用”是驱动基本概念衍生的重要关系。今后，我们希望能透过更多其他类别意符的分析，以观察不同类别意符在知识结构上的特性与共性，以期能更系统化地建立出汉字意符的知识结构，并提供汉字驱动知识工程与语意运算的基础架构。

引用文献

- Chen, S.-Y. (陈圣怡), C.-R. Huang (黄居仁), and Y.-J. Yang (杨雅君). 2008. The knowledge system of radicals: A study on the representation and cultural implications of yu4 (jade) and shi2 (rock). Presented at The 4th International Conference on Literature and Information Technology (ICLIT 2008). Hong Kong: City University.
- Chou, Y.-M. (周亚民) and C.-R. Huang. 2010. Hantology: Conceptual system discovery based on orthographic convention. In C.-R. Huang, et al, eds., 2010a. Pp. 122 - 43.
- Huang, C.-R., N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari and L. Prévot, eds. 2010a. *Ontology and the Lexicon: A Natural Language Processing Perspective*. Cambridge: Cambridge University Press.
- Huang, C.-R., R.-Y. Chang (张如莹), and S.-B. Li (李祥宾). 2004. Sinica BOW (Bilingual Ontological Wordnet): Integration of Bilingual WordNet and SUMO. In *Proceedings of the 4th LREC*. Lisbon, Portugal.
- . 2010b. Sinica BOW: A bilingual ontological wordnet. In C.-R. Huang, et al., eds., 2010a. Pp. 201 - 11. Cambridge: Cambridge University Press.
- Huang, C.-R., S.-Y. Chen, S.-K. Hsieh (谢舒凯), Y.-M. Chou, and T.-Y. Kuo. 2008a. Linguistically conventionalized ontology of four artifact domains: A study base on Chinese radicals. Presented at the Workshop on Linguistic Studies of Ontology, the 18th International Congress of Linguists. Seoul, Korea.
- Huang, C.-R., Y.-J. Yang, and S.-Y. Chen. 2008b. An ontology of Chinese radicals: Concept derivation and knowledge representation based on the semantic symbols of four hoofed-mammals. *Proceedings of the 22nd Pacific Asia Conference on Language, Information and Computation (PACLIC2008)*. Pp. 189 - 96.
- Hsieh, S.-K. 2006. Hanzi, concept and computation: A preliminary survey of Chinese char-

- acters as a knowledge resource in NLP. Ph. D diss. , Universit t T bingen , Germany.
- Niles , I. and A. Pease. 2001. Toward a standard upper ontology. In *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems*. Ogunquit , Maine.
- . 2003. Linking lexicons and ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In *Proceedings of the IEEE International Conference on Information and Knowledge Engineering (IKE 2003)* . Las Vegas , Nevada. Pp.412 –6.
- Pustejovsky , J. 1995. *The Generative Lexicon*. Cambridge , MA: The MIT Press.
- Wong , S.-H. S. and K. Pala. 2002. Chinese characters and top ontology in EuroWordNet. In *Proceedings of the First Global WordNet Conference 2002*. Mysore University , Karnataka , India.
- 洪嘉馥、黄居仁，2012，以汉字知识本体为出发的汉字教学系统。第十三届汉语词汇语义学研讨会 (CLSW 2012) 论文集，武汉大学，湖北。461 –6 页。
- 黄居仁，2009，从词汇看认知：词汇语义学研究的趣味。见苏以文、毕永峨主编，《语言与认知》。台北：台大出版中心。203 –28 页。
- 黄居仁、陈圣怡、周亚民，2010a，语言的知识与知识的语言：由《说文解字》出发的知识本体研究。见潘悟云、沈钟伟主编《研究之乐：庆祝王士元先生七十五寿辰学术论文集》。上海：上海教育出版社。106 –22 页。
- 黄居仁、谢舒凯、洪嘉馥、陈韵竹、苏依莉、陈永祥、黄胜伟，2010b，中文词汇网络：跨语言知识处理基础架构的设计理念与实践。《中文信息学报》第 2 期，14 –23 页。
- 许 慎撰，崔枢华、何宗慧编，2000 [121]，《说文解字》。北京：北京师范大学出版社。
- 周亚民，2005，汉字知识本体——以字为本的知识结构与其应用示例。台湾大学博士论文。
- ，2009，汉字知识本体在汉字教学的应用。《华文语文教学研究》第 1 期，91 –112 页。
- 周亚民、黄居仁，2005，汉字意符知识结构的建立。第六届词汇语义学研讨会，厦门。
- ，2006，汉语文字和词汇知识在计算器的表达——历史变迁的观点。见何大安等编辑，《山高水长：丁邦新先生七秩寿庆论文集》。南港：中研院。595 –611 页。
- ，2013，汉字知识的形式表达。《当代语言学》第 2 期“知识本体”专号，142 –61 页。

第一作者简介

见本期第 293 页。

- 作者通讯地址： 黄居仁 Huang Chu-Ren , Faculty of Humanities , The Hong Kong Polytechnic University , Kowloon , Hong Kong
- 洪嘉馥 台湾台北市 和平东路一段 162 号 台湾师范大学华语文与科技研究中心(普 317 –1)
- E-mail: churen. huang@ inet. polyu. edu. hk(黄居仁)
 jiafeihong@ gmail. com (洪嘉馥)
 sheng. eagles@ gmail. com (陈圣怡)
 milesymchou@ gmail. com (周亚民)

technology critical to its development is ontology. This paper presents an overview of the development of ontologies and its important role in language processing. It also introduces various interdisciplinary studies on ontologies and language sciences. In particular, it discusses how ontologies address the challenges that Chinese language processing faces in the Semantic Web. It also summarizes the 10 papers in the Special Issue (Vol. 15 , No.2) and the Special Section in this issue (Vol. 15 , No.3) focusing on ontology.

Keywords: Semantic Web , ontology , Chinese language processing

Chu-Ren HUANG , Jia-Fei HONG , Sheng-Yi CHEN and Ya-Ming CHOU , Exploring e-vent structures in Hanzi radicals: An ontology-based approach

Shuowen Jiezi (《说文解字》) is organized according to the radical forms as semantic symbols. Characters are classified according to radicals, and their meanings cluster around the basic concept of the semantic symbol. Therefore, in this paper, we assume that *Shuowen Jiezi* radicals can reflect the conventionalized conceptualization when Chinese character orthography was invented. According to our analysis elaborating Generative Lexicon Theory by Pustejovsky (1995), we found that the ontology expressed by Hanzi radicals has already had the strong conceptual derivation and knowledge reasoning ability as described in the Generative Lexicon. In this study, we take as our research objects Chinese radical 艹 *cao3* representing grass, and the radicals representing Five Sense Faculties in *Shuowen Jiezi*, namely 目 *mu4* (“eye”), 耳 *er3* (“ear”), 口 *kou3* (“mouth”), 鼻 *bi2* (“nose”), and 舌 *she2* (“tongue”) which all belong to “body part” class in SUMO concepts. In addition, we assume that semantic symbols represent basic concepts, and identify the semantic relation between each derived character and its basic concept to construct a conventionalized ontology headed by that concept. Finally, we contrast the semantic symbol generative structure for the Five Sense Faculties with the modern senses of these single-character words in the Chinese WordNet. It is observed that the same set of derivational relations applies.

Keywords: Chinese radicals , ontology , Hanzi , lexical semantics

SU Qi , Evidentiality and other linguistic evidence in computational language processing

As a distinct linguistic category, evidentiality plays an important role in statement and communication. However, evidentiality is complicated in that it refers not only to grammatical categories but also to lexical items as well as specific expressions. Generally speaking, certain lexical items and specific expressions can be used to represent evidentials. Many linguists are committed to the study of their types and degrees of reliability reflected. This paper aims at rounding up the findings of evidentiality based on the evidence drawn from annotated English and Chinese corpora. It can be seen that there are some patterns in defining evidentials so as to evaluate the reliability of information. Also reported are the findings of other linguistic cues which may affect people’s judgment of information reliability as well.

Keywords: evidentiality , evidentials , reliability , annotated corpus