《当代语言学》第 15 卷 2013 年第 3 期 284 – 293 页，北京

知识的系统与知识系统的建构：知识本体语言科学整合研究

黄居仁　李逸薇　香港理工大学

提要　语意网络被形容为互联网的下一代，而它所需要的关键技术正是知识本体。本文概述知识本体的发展及其在语言处理中所扮演的重要角色，介绍知识本体和语言科学的跨学科研究，重点讨论知识本体如何解决中文语言处理在语意网络中所面临的挑战。本文还介绍了作者编审的“知识本体”专题中 10 篇论文的主要内容。

关键词　语意网络　知识本体　中文语言处理

1. 知识的表达系统：知识本体

语言是用于承载与交流知识与信息的表达系统。而知识本体(ontology) 则是用来描述一个系统内部知识体系的架构。虽然知识本体在哲学领域具有较长远的研究，但是近些年来才渐渐受到其他领域，特别是信息科学领域的重视。其主要原因是知识科技与知识经济的蓬勃发展，同时也或多或少受到“语意网络”(semantic web) 兴起的影响。Berners-Lee 等(2001) 在《科学美国人》中宣告语意网络的愿景是：电脑能自动“阅读”并理解网路资料的内容。这种自动阅读语意的能力，正是需要知识本体去帮忙理解每个网页，以表示网页中知识的架构与内容。因此，如何建立跨语言的共享知识本体架构用以正确有效地表达不同领域知识，成为相关学术领域非常重要的研究议题。

语言的知识系统，就是指人类的知识系统。人类所传递的信息真是因为语言所赋予的结构才成为知识。因此，所谓的“知识本体”并非用于表达细节知识，而是用于描述知识的概念结构。一般来说，知识本体的架构越抽象、层次越高，其所涵盖的知识范围也越广；但是，相对而言，这样的知识本体距离实用的知识层次越远，能表达的实际知识越少。因为语言涵盖了所有人类所能表达的知识，所以可以认为任何语言都有其内涵的知识本体，所有讲该语言的人，都不自觉地在使用这个隐含的约定俗成系统。从语言到文字，又是更进一步的约定俗成。在众多类型的文字中，汉字是一种独特、具外显表义功能的书写系统。其不仅跨越三千年历史，而且为汉字文化圈中不同语言所采用。因此，使用汉字语言所构建的隐含知识表达系统，可以说是全世界使用人口最

多、最稳定、表达知识最丰富的知识本体。

语言的主要功能是传递与表达讯息。当这些讯息经过系统转化为知识时，语言也自然成为承载知识的主要媒介。在知识本体的研究中，我们发现承载知识的表达系统本身必然具有严谨的知识体系。换言之，语言能作为有效沟通工具的原因在于使用同一语言的人必须同时接受该语言内涵的知识系统。在此前提下，语言之间知识系统的接口变得十分重要。该研究点正是黄居仁等(Huang，et al. 2010)一书所代表的新研究方向。书中将该研究内容称为“本体-词汇界面”(OntoLex，即 Ontology-Lexicon interface)。该研究内容的核心议题是如何发现并运用语言中约定俗成的知识系统。

简言之，从知识表达体系的观点出发；语言代表了约定俗成且表达能力强的多样化体系；语言具鲁棒性(robustness)、能容许不完全规范的表达形式；但多样体系间的信息不易交换共享。相对而言，知识本体则是逻辑规范化、内部结构严谨的系统。其设计目标，便是不同知识系统间共同的表达与信息交换系统。知识本体对非规范形式的容许度低；用以换取不同知识体系间信息融合与交换的有效平台。如何在知识本体研究中引入语言的多样性与文化表达力；如何用知识本体的严谨知识表达架构，对语言间的通用性与差异性提供基于概念架构的解释，正是时下研究的主要议题。可以同时从知识本体的角度出发，对汉语信息处理做出详细的研究和探索。

2. 知识本体：上层知识本体与特定知识本体

知识本体可针对的应用范围分为上层知识本体与特定知识本体。上层知识本体可以是任一完整知识系统的表达，可能涵盖特定学术领域(如化学知识领域、生态学知识本体等)，应用领域(如环保知识本体、旅游知识本体等)，或特定的机构(如大网购公司都有其内部知识本体)。特定知识本体的主要功能是确保特定领域内知识表达的一致性，以消除误会并促进信息共享的效率。特定知识本体，包括了其专有词汇集、词汇定义及词汇/概念间的关系架构等。只针对单一语言发展的知识本体，在个别语言可有受其文化历史影响的知识系统，也可视为特定知识本体。特定知识本体建构有两个重要基本要求：内部知识表达的完整与一致，以及知识表达架构与其他知识系统，特别是上层知识本体的兼容性。

以颜色为例。不同文化语言间，对颜色的分类与命名，有不同的系统(如中文原来无“绿色”一词，只有蓝，青等相近概念)；不同领域对颜色定义方式不同(物理学以波长，色彩学以原色组合等)；应用上又可能因颜料与表面材质需要调整。因此，不同领域与应用范围，都需要有完整的内部有关颜色概念的定义与表达。“景泰蓝”、“哥本哈根蓝”、“宝蓝”(royal blue)，都需要有文

化/领域特定的精确定义，却有跨领域概念(瓷器、服饰、汽车、文学等)共享的需求。因此，除了特定知识本体的完整描述与表达，还需要有能融通各特定知识本体的共同知识表达架构。

不同知识系统对各种概念与名词具有不同的定义与定位，往往无法完全兼容。因此，为了知识系统之间的沟通，有必要构建一个共有的上层知识系统。于是，美国电气电子工程师学会(IEEE)成立了一个工作小组，名为"标准上层知识本体工作小组"，并建立了一个"建议上层共享知识本体"(Suggested Upper Merged Ontology，简称 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。SUMO 知识本体，目前由 Articulate Software 负责维护发展，重要资料可经由网站[①]取得。

SUMO 上层共享知识本体目前有英、中、德、意、捷、印六种语言版本可供选择，主要内容包括概念及定理(axioms)。上层知识本体的内容被限制在后设(meta)的范围中，即包括了一般、抽象或者哲学的概念，以便能涵盖范围较广的领域。特殊领域具体的概念则不被包括在上层知识本体中。此类型的知识本体可以帮助特定领域(例如：军事、医药、财政等等)的知识本体结构的快速建立与交换。虽然特定领域知识本体中的细部知识往往同其他知识本体有所出入，但是由这些知识向上推至的较高概念层次，即 SUMO 表达的层次，必定有一致的分类。SUMO 的作用就是提供具有一致性的高层概念结构，所有的知识本体的高层次概念都可以对应到 SUMO 中的概念。具体来讲，SUMO 中的概念组成结构是将概念点以承继树的方式做连接，一共包含 11 个大类、3912 个概念。此外，SUMO 的另外一个特色在于其关键设计能够与英语词汇网路的连结(Pease and Fellbaum 2010)。这一特性正好满足了人类跨领域语言使用的特性，即能够跨越领域障碍表达所有知识的本体架构。目前，SUMO 已经同英语词汇网路 WordNet 1.6 及 2.0 版本构建连接，使得任何领域的知识，都可以藉词汇的连接，建立正确的知识本体位置。完整的 SUMO 资料库，可以在其官方网页[②]检索或下载。

目前学术界通用的知识本体架构，除了 SUMO 外，DOLCE[③] 是另外一个广受关注的本体框架。DOLCE(Gangemi，et al. 2002)并非一个现成的通用标准知识本体。相反，它的目的是作为出发点用于同其他未来模块进行比较和描述，同时也用于澄清现有知识本体(例如 WordNet)背后隐藏的假设。与其他知

① http://www.ontologyportal.org/ [accessed 1，May 2013]
② http://www.ontologyportal.org/ [accessed 1，May 2013]
③ http://www.loa.istc.cnr.it/Papers/D18.pdf [accessed 1，May 2013]

识本体不同的地方在于，DOLCE 的目的是捕捉关于自然语言与人类常识的本体论范畴。然而，DOLCE 并不意味着严格参考形而上学式的世界内在本质。相反，其引入的类别取决于人的感知，文化的烙印和社会习俗。在此意义上，他们只专注在描述性的概念上，而这些概念是建立在现有形成的概念基础上。DOLCE 是一种特定的知识本体，所以所牵涉的概念会被限制在某个特定范畴内。当然，通用性在组织和具体化概念属性的时候也充分被考虑，不过这种考虑取决于概念属性本身组合和具体化的要求。

3. 知识本体与汉语理论与计算语言学研究

从语言学理论与计算研究的观点出发，知识本体理论的发展带出了以下几个重要的议题：(1) 各语言间是否有共同的概念系统？(2) 个别语言的语意与概念系统能否以知识本体这样的逻辑体系描述？(3) 语言的创造性可否有系统性的描述与解释？(4) 知识本体如何解决语意计算与概念理解的问题？④

世界各语言间是否有共同的概念系统？以及语言间如存在系统性的概念差异，这些概念系统差异是否能以知识本体描述？知识本体在这个研究方向的最重要贡献，是提供了一个没有文化偏见的共同描述系统，作为比较研究的规范架构。Prévot 等(2006)，Huang (2007)，Huang 等(2007) 提出了基于斯瓦迪司词表(Swadesh List) 基本词汇的研究方法与初步结果。此外，语言描述共同知识本体 (General Ontology for Linguistic Description，GOLD) ⑤则为语言学与语言典藏的知识内容提供了共同的知识表达架构。Schalley 和 Zaefferer (2007)、Huang 等 (2010) 两本重要选辑也各收入了一些重要的研究论文。

个别语言的语意与概念系统能否以知识本体的逻辑体系描述？这个研究议题和上个议题具有明显不同的地方：在第一个议题中，知识本体是作为描述与分析语言的工具；而在本研究议题上，则直接把语言当成(约定俗成的) 知识系统。因此，该议题研究的基本问题，是人类语言的知识系统与逻辑的知识系统有何异同？这个研究方向的主流研究方法是以词汇网路(WordNet，Fellbaum 1997) 为基础的研究 (Pease and Fellbaum 2010)。词汇网路(简称词网，WN) 所有同义词的集合(同义词集，synset) 为概念节点；建立语言的概念架构，并以上下位词关系作为架构的骨干；辅以其他各种词义关系，建立个别语言的完整概念架构。故词网常被称为“语言知识本体”。词网到规范知识本体间的对应，也是计算语言学与知识本体研究中最受注重的研究方向之一。研究的基本议题便是本体-词汇界面(参见 Huang，et al. 2010)；本刊 2013 年第 2 期“知

④ 第四个问题为目前计算语言学与自然语言处理最关心的问题。

⑤ http://linguistics-ontology.org/ [accessed 1，May 2013]

识本体”专号中的论文，大部分都涉及了这个议题的研究。中文知识本体的相关研究，必须由完整的中文词网作为基础。中文词网相关研究相当丰富；但是，这些研究中能真正整合知识本体研究的，目前只有黄居仁等(Huang，et al. 2010)的中文词网(Chinese WordNet⑥)。

语言作为知识系统的概念带入后，语言的创造性可否有系统性的描述与解释便成了以知识本体进行语言学研究的一个新研究方向。这里提到的语言创造性，包括了隐喻、语意延伸、幽默、双关语、反讽等非字面意义的语言用法。这些语言创造性用法的共同之处，是把语言表达的字面本意引导到新的方面，赋予了新的诠释。在传统语言学理论的分析中，这些创造性是新奇而不能以规律解释的。但是，在引进知识本体的概念后，语言创造性的丰富内容，及其听话者的会心理解能够得到一个较好的解释。语言创造性的产生，并非是随性或任意的。语言的创造性，其实是在语言原有的知识系统上，做整个系统的转移与对应。因此，说话者可以继续延伸发挥；听话者也可以举一反三，并利用原有的语言知识理解新的创造性说法。Huang 等(2006)开启了这个研究方向的论文；紧接着，钟晓芳的博士论文(Chung 2007)及后续的相关研究(包括本刊2013 年第 2 期“知识本体”专号中的论文，详见下文)探讨了与隐喻研究相关的议题。

在计算语言学研究领域中，语意研究这个核心议题一直被认为是很不容易处理的。主要的原因，应该是缺乏一个表达能力强、规范化，并具有效计算处理工具的概念与语意架构。知识本体的提出，正好填补了这个空缺，也促进了最近几年来语意与概念自然语言处理研究的蓬勃发展。在知识系统的架构概念上，计算语言学近来的两个研究方向，很值得理论语言学的参考。第一个研究方向是信息品质的研究。语言与文本传达信息，把文字转换到信息内容是理解的第一步，但解读信息内容后，更重要的一步，是判定信息内容是否可靠，有多少价值。这样的信息品质评价，人们在听话时几乎随时都在处理。但对计算机处理而言，由解读信息内容到对信息的品质评价，是为信息内容加值，最具挑战性，也是最关键的一个环节。信息内容的评价，当然也不能与信息所表达的情感与情绪分割。语言学理论，目前为止在这两个方面着墨甚少。语言的情感表达，与以语言为本的信息品质分析，是语言学研究亟待发展的两个方向。计算语言学领域针对知识本体的研究，已有相当蓬勃的发展。也有董振东开发的知网(HowNet)的研究传统。不过知网的研究集中在语言处理与机器翻译方向。

汉语可不可以用来表达逻辑严谨的知识本体？该问题牵涉到本体研究的实

⑥ http://cwn.ling.sinica.edu.tw/ [accessed 1，May 2013]

际发展。以语意为出发的知识表达与检索是不可避免的大趋势，而以“知识本体”来描述知识内容与概念架构，也几乎成为知识使用的必然手段。虽然每种语言与文化具有各自的知识与概念内容，也在各自的文化典藏中得到表现。但是，在知识的传播与共享时，这些知识内容必须具有共通的、可转换的架构。因此，汉语信息在构建知识本体方面的挑战，不仅仅是要提出一个合理且有足够涵盖的知识本体用于描述汉语内容的知识架构，而且要保证这个架构能够与其他语言(的知识本体)相互转换及沟通。

其实，在回答上述问题前，我们还需要讨论一下语意网的基本问题：是否有一个共享的知识本体，可以完善表达世界上所有的知识，从而可以作为所有知识交换的标准？由于有许多彼此矛盾的知识体系(比如说各种宗教、不同主义)，全世界的知识当然不能用一个单一架构表达。但是，从知识工程的角度出发，是有可能把最上层的知识概念表达出来，建立成一个共同的架构。这就是 SUMO 的由来。SUMO 是由国际电机工程师学院 IEEE 标准上层知识本体工作小组所建置，共有约一千个概念组成知识本体结构。上层的知识本体限制在“后设”的概念、也就是一般、抽象或者哲学概念。所有概念足够涵盖广阔范围的领域区域最上层的知识结构。特定领域具体的概念不被包括在上层知识本体中，但是这样的知识本体确可帮助特殊领域(例如：医药、财政、项目等等)的知识本体结构的建立。SUMO 希望藉由最高层次的知识本体，鼓励其他特殊领域知识本体以其为基础衍生出其他特定领域的知识本体，并为一般多用途的术语提供定义。目前，SUMO 已经和英语词汇网络 WordNet 1.6 版本作初步的连接，也就是说，可以由任一英语词汇出发，得到相对的知识概念节点。SUMO 的出现，使得在信息应用上，有一个可以用来表达所有知识的共同知识本体架构。

SUMO 的初期工作主要集中在英文语言中。针对汉语，台湾中研院语言所中文词网小组最先依据 SUMO 2002 年版的资料进行系统界面及概念节点的中文化，其内容主要参考 Wordnet 1.6 的英中对译。紧接着，中研院语言所于 2004 年对 termformat 及 format 的汉语进行了翻译和修正，并于 2006 年完成中层知识本体 MILO (MIdLevel Ontology)的中文化。中研院语言所的同仁从汉语的词出发，将 SUMO 以及中英对译的词汇网络结合，建成了“中央研究院双语知识本体词网”(The Academia Sinica Bilingual Ontological WordNet⑦，Huang，et al. 2004，2008)，简称“研究院知识词网”(Sinica BOW)。他们的愿景，就是在这个知识库的平台上，逐渐建立可以跨越不同语言与知识系统鸿沟的工具。中研院的双语本体知识词网，同时有中英双语互查，以及由任一个语言检

⑦ http://bow.sinica.edu.tw/ [accessed 1，May 2013]

索知识本体的功能。也就是说，可以由任何一个汉语或英文词汇的词义，查到在 SUMO 的概念架构上属于那个词汇的概念节点。因此，可以认为该本体已经提供了由语言到知识架构的接口。此外，在语言学习上，该本体可以帮助建立以知识体系及相关概念为基础的学习系统。综上所述，对汉语是否能提供严谨知识表达架构这个语意网上的关键问题，“中研院双语知识本体词网”已提供了基本肯定的答案。

4. 专文内容与议题简述

本体构建旨在对知识体系的概念和关系建模并形成体系化知识，从而辅助计算机进行智能化的处理。这种方式可以理解为知识本体的自然语言处理的相关研究。该研究面临的主要问题在于如何发掘、交换、并产生新知识。因为上位本体包含的概念是与领域无关的通用概念集，所以这些本体知识已经基本在已有的框架中得到充分体现。因此，目前面向知识本体的自然语言处理研究集中在通过计算机算法自动提取而获得的本体主要用于特定领域的概念知识。

此次由黄居仁、李逸薇编审的“知识本体”专题共 11 篇文章，在本刊分两期发表：2013 年第 15 卷第二期知识本体专号(8 篇)；第三期知识本体研究专栏(3 篇，含本篇)。

本刊“知识本体专号”开篇“自然语言语义、语义自动化处理与知识本体——写在知识本体专号前面的话”一文，指出了目前计算器语言学所面对的问题，从而引出知识本体在中文语言处理中的必要性。

专号中陆勤等“‘自下而上’与‘自上而下’本体构建方法的探讨”一文，利用英文的上位本体 SUMO，通过自上而下的方法建立中位本体。此方法充分利用中英词典和英文的词汇语义知识 WordNet 对上位本体的映射关系，并通过计算机算法利用知识和领域词汇的统计特征进行两个阶段的消歧来完成。这种方法着重于与上位本体进行对齐从而进行下位本体的扩展。因此，该方法很好地承袭了上位本体的现有构造。基于下位本体的特性，他们还进一步介绍了一种自下而上的本体构建方法。此方法更适用于发现新的领域词汇并抽取其相关的领域核心概念，进而对已有的领域本体进行扩充。同时，该方法还可以应用本体建构提供可衔接的本体知识，有利于维护知识空间的完整和关联。

汉语语言内涵丰富的知识内容，可在知识为本的信息处理中得到发挥。其中，汉字的知识本体和形式表达是相当重要的研究内容。专号中周亚民和黄居仁的“汉字知识的形式表达”一文，说明如何在计算机建立汉字知识，以及如何用形式语言表达汉字知识。与已有的汉字资料库不同的是，他们以语意网的形式语言描述汉字知识。具体包括：字形内外结构、意符与声符、字义与衍生词等。这些汉字知识皆与上层知识本体(SUMO)对应，作为汉字知识的上层知

识。这样有助于汉字知识与其它知识本体分享知识，从而对汉语研究有所开展和启发。本期中黄居仁等的“汉字所表达的知识系统: 意符为基本概念导向的事件结构”一文，则以《说文解字》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。此研究厘清《说文》“艹”、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“艹”、“五官类”基本概念带领的个别意符知识系统。该研究显示以“艹”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。

近年来，建构词汇知识资源(如词网) 已经成为词汇语意学与本体知识工程的共同关注的焦点之一。其中，语意关系的标记是建构系统不可或缺的部份，在语言与认知研究上也具有深刻的意涵。在本专题中，有两章内容关注知识本体处理汉语词义分类与预测方面的研究。专号中谢舒凯的“汉语词汇语意关系: 分类、逻辑检证与知识本体评测方法初探”一文，讨论了词汇语意关系的研究现况，特别聚焦在所谓的概念包含关系的自动分类、发掘与评测方法。该文还提出利用一种形式知识本体评测的方法论来进行关系评价，以期达到逻辑上面的严谨，从而提高词网的质量。

此外，专号中林素朱等的“广义知网的词汇知识架构与语义表达”一文，则以自然语言理解为目标，提出“广义知网”架构。该语意表达框架有效地应用于语义的合成和分解上。在此架构中词汇的语义皆以事物义原、基本概念及其间的关联来定义，实词和虚词的语义，皆有一致的表达方式，词汇语义定义时，义原之间的关联能完整而直接的呈现，从而进一步使得广义知网具有语义合成与分解的机制，短语或句语义可以用词汇语义的定义式来合成。并且所产生的词汇与词组语义表达式，会有近乎标准且唯一的表达形式，以达到语义自动了解的第一步。

非字面义的表达，例如隐喻与言据性，在陈述与交流中具有重要地位。其复杂的结构及抽象的含义，通常涉及到词汇语义、句法及语用等诸多方面的分析。这些非字面义的表达一直是信息处理一个重要的障碍。专号中钟晓芳等的“中英知识本体和概念隐喻”一文，以中英文为例透过上层知识本体(SUMO) 剖析了不同文化使用的概念隐喻的异同。此方法亦解释了汉英学习者学习某些概念隐喻的困难之处。本期中苏祺的“计算机语言处理中的言据性及相关语言学线索”一文，深入探讨汉英之间的言据性表达的联系和差别。在分析标注语料库的基础上，总结常见言据的类型与问题，以及除了基本词汇据素之外的与言据性表达相关的语言线索。

专号的最后两篇文章介绍了如何使用知识本体帮助自然语言处理问题。顾

曰国和张永伟的“静态图像库的信息检索与知识本体”一文探讨如何利用知识本体处理计算机静态图像。文中提出了两组六层的处理模型，成功解决了计算机处理图像时的调用问题，并将人类知识进行了形式化。李宏等的“简单本体在实用信息抽取中的使用及针对实用本体的高级信息抽取”一文中，设计了一个信息抽取的本体模型，以及一个关系抽取的框架系统——DARE。DARE 使用了机器学习方法，可以自动学习关系抽取的语言模板，并应用这些模板来抽取关系实例。值得一提的是，DARE 系统已被深度应用于英语文本的关系抽取中。李宏等人也进一步使用 DARE 来处理汉语新闻文本，从中学习语言模板和抽取关系实例，并与英文处理的结果进行了比较研究。

5. 结论及展望

语言的本质是知识系统。语言学研究的议题包括：如何习得该知识系统；如何将知识以约定俗成的符号系统表达；如何利用符号的结合方式与顺序来表达复杂知识关系，知识系统如何表达文化内涵；如何在不同知识系统间交换知识并化除误解等。然而，过去的理论与计算语言学研究，建立在符号系统与结构为基本的理论架构上，无法直接描述或处理知识的系统性与知识表达的架构。值得庆幸的是，现有的关于知识本体理论的研究，打通了这个瓶颈。“知识本体”专号收集了结合知识本体的汉语语言科学研究初步成绩。虽然本专题所呈现的是初步的尝试，主要是要建立相关研究的基础架构及探讨相关方法，但已明确显示了在知识系统出发的研究方面所拓展的丰富研究议题，并提供了深具解释性与启发性的研究结果。汉语知识本体的研究方兴未艾。我们希望，回到知识系统与语意层面的研究，能提升汉语语言学研究的国际能见度，使国内语言学与计算语言学研究能登国际舞台的大雅之堂。

引用文献

Berners-Lee，T.，J. Hendler and O. Lassila. The Semantic Web. *Scientific American*，May 2001：29－37. 中译本：2002，高虹译，黄居仁审，语意网（Semantic Web）。《科学人》2002 年 5 月号。

Chung，S.-F.（钟晓芳）. 2007. A Corpus-driven approach to source domain determination. Ph. D. diss.，Graduate Institute of Linguistics，National Taiwan University，Taipei.

Fellbaum，C.，ed. 1997. *Wordnet*：*An Electronic Lexical Database*. Cambridge，MA：The MIT Press.

Gangemi，A.，N. Guarino，C. Masolo，A. Oltramari，and L. Schneider. 2002. Sweetening ontologies with DOLCE. *EKAW* 2002. Pp. 166－81.

Huang，C.-R.（黄居仁）. 2007. Towards a common conceptual framework of language documentation. In *Proceedings of the International Conference on Endangered Austronesian Language Documentation*. Providence University，Taichung.

Huang, C.-R., L. Prévot, I.-L. Su (苏依莉), and J.-F. Hong (洪嘉馡). 2007. Towards a conceptual core for multicultural processing: A multilingual ontology based on the Swadesh list. In T. Ishida, et al., eds., *Intercultural Collaboration*. Invited and Selected Papers from IWIC 2007 Kyoto. Lecture Notes in Computer Science. Vol. 4568. Pp. 17 – 30.

Huang, C.-R., N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari and L. Prévot, eds. 2010. *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.

———. 2008. Sinica BOW: A bilingual ontological wordnet. In C.-R. Huang, et al. eds., *Ontologies and Lexical Resources for Natural Language Processing*. Cambridge Studies in Natural Language Processing. Cambridge: Cambridge University Press.

Huang, C.-R., S.-F. Chung and K. Ahrens. 2006. An ontology-based exploration of knowledge systems for metaphor. In R. Kishore, R. Ramesh, and R. Sharman, eds. *Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems*. Berlin: Springer. Pp. 489 – 517.

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.

Pease, A. and C. Fellbaum. 2010. Formal ontology as interlingua: The SUMO and WordNet Linking Project and Global WordNet. In C.-R. Huang, et al., eds. 2010. Cambridge: Cambridge University Press. Pp. 31 – 45.

Prévot, L., C.-R. Huang, and I-Li Su. 2006. Using the Swadesh list for creating a simple common taxonomy. In *Proceedings of the 20th Pacific Asia Conference on Language, Information and Computation*. Wuhan, China.

Schalley, A. and D. Zaefferer. 2007. *Ontolinguistics: How Ontological Status Shapes the Linguistic Coding of Concepts*. (Trends in Linguistics. Studies and Monographs 176.) Berlin/New York: Mouton de Gruyter.

黄居仁、谢舒凯、洪嘉馡、陈韵竹、苏依莉、陈永祥、黄胜伟，2010，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。

第一作者简介

黄居仁，男，博士，香港理工大学中文研究与应用讲座教授。研究兴趣：词汇语义学、计算语言学、语料库语言学、知识本体与语言资源。代表作： *Ontology and the Lexicon* **(共同主编) 和** *Computational Linguistics and Beyond* **(共同主编) 等。电子邮件：** churen. huang@ inet. polyu. edu. hk

HUANG Chu-Ren. male, Ph. D., is Chair Professor of Applied Chinese Language Studies at The Hong Kong Polytechnic University. His research interest includes lexical semantics, corpus linguistics, computational linguistics, ontology, and language resources. His major publications are: *Ontology and the Lexicon* (co-ed.) and *Computational Linguistics and Beyond* (co-ed.). E-mail: churen. huang@ inet. polyu. edu. hk

作者通讯地址： **黄居仁** Huang Chu-Ren, Faculty of Humanities, The Hong Kong Polytechnic University, Kowloon, Hong Kong

李逸薇 Sophia Yat Mei Lee, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Kowloon, Hong Kong

E-mail: sophiaym@ gmail. com(**李逸薇**)

《当代语言学》第 15 卷 2013 年第 3 期 376 – 378 页，北京

Abstracts of Articles

SHEN Jiaxuan and YUE Yao，Explaining experiment results on word classes: Toward an updated grammatical theory

The results of the experimental studies on word classes are different from and even contrary to each other. It is either because of the employment of different experiment methods and tasks，or because of the theory of word classes by which we explain the experiment results. In this article，two groups of experiments relevant to Chinese word classes study are reviewed and examined，one being perceptual experiments on infants or children，and the other fMRI experiments on adults. It is proposed that theory of word classes needs to be updated in order to make a reasonable explanation and draw a correct conclusion on the experiment results. In terms of the noun and verb distinction，two types of distinctive patterns should be established，namely 'noun-verb dissociation pattern' and 'verb as a subcategory of noun pattern'. Not only can grammatical theory be used to explain experiment results and be tested by experimental studies，but also experimental studies call for updating the grammatical theory which will in reverse improve and deepen experimental studies.

Keywords: word classes，perceptual experiment on infants and children，fMRI，noun，verb

YANG Suying and HUANG Yueyuan，A survey of the four major Chinese aspect markers in different modes of discourse

This paper reports the results of a corpus-based study on the distribution of the four major Chinese aspect markers 了$_1$ *le*$_1$，着 *zhe*，在 *zai* and 过 *guo* in different modes of discourse including conversation，fiction and news. We have made some new findings and have also found corpus data support for some previous findings. Our findings include: *le*$_1$ is the most frequent of the four aspect markers in all the three modes of discourse，while *zai* is the least frequent. *Zhe* and *zai* do not appear in narrative clauses. More occurrences of *guo* appear in non-narrative clauses but there are also some occurrences that appear in narrative clauses. Most occurrences of *le*$_1$ appear in narrative clauses. An endpoint and completion or termination of an event are the necessary requirements for the appearance of *le*$_1$. The appearance of *le*$_1$ is also conditioned by the story line，the prosody of the clause，the situation type of the clause，and the type of complements after the verb.

Keywords: distribution of aspect markers，modes of discourse，discourse structure

Chu-Ren HUANG and Sophia Yat Mei LEE，Knowledge systems and the construction of knowledge systems: Introducing interdisciplinary studies on ontologies and language sciences

The Semantic Web has been promoted as the future generation of the World Wide Web. A

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 event 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** *cao*3 representing grass, and the radicals representing Five Sense Faculties in *Shuowen Jiezi*, namely **目** *mu*4 ("eye"), **耳** *er*3 ("ear"), **口** *kou*3 ("mouth"), **鼻** *bi*2 ("nose"), and **舌** *she*2 ("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