文章编号：1003-0077（2011）00-0000-00

语言同现网、句法网、语义网的构建与比较*

赵怿怡[1]，刘海涛[2]

（1.厦门大学，福建省 厦门市 361005；2. 浙江大学，浙江省 杭州市 310058）

摘要：网络方法应用于语言研究是语言研究大数据时代的新趋势。语言是一个多层级的符号系统，选择哪种语言单位作为网络节点，选择哪种语言单位间的关系作为网络联结，影响到语言网络的结构和功能。该文梳理了以汉语词为单位，以同现、句法、语义关系为联结依据的几类网络构造方法，并针对同一文本构造三类网络发现：句法网络的网络直径、平均路径长度远小于同现网络，实词在语义网络中占据中心节点位置。这提示我们网络分析方法的应用仍要以可靠的语言学理论为指导，从语言学内部出发才能更好解释各类语言网络的差异。

关键词：同现网； 句法网； 语义网

中图分类号：TP391 **文献标识码：**A

The Structure and Comparison among Language Co-occurrence, Syntactic, Semantic Network

Zhao Yiyi[1], Liu Haitao[2]

(1. Xiamen University, Xiamen, Fujian Province 361005, China; 2. Zhejiang University, Hangzhou, Zhejiang Province 310058, China)

Abstract: Network method applied in language studies is a new trend in the age of big data. Language is a multi-level system of symbols. Language networks based on different language units and the relationship have special features. We constructed word co-occurrence network (on the basis of the adjacency of words), syntactic network (on the basis of syntactic theory-dependency grammar) and semantic network (on the basis of conceptual relation) for the same text. We find that syntactic network diameter, average path length is much smaller than co-occurrence network, and content words in the semantic network occupy central node locations. This suggests that we had better apply the network analysis based on linguistic theory, which will contribute to better explain the differences of various language networks.

Key words: co-occurrence network; syntactic network; semantic network

1 引言

语言是一种复杂动态系统[1-5]。它在各个层级表现出高度的复杂网络结构（语音，词汇，句法，语义）[6-7]。此类结构的形成与演化是数百万使用者长期使用的结果，使用者适应并改变语言使它满足当下交流的需要[8]。语言本身和语言所反映的人类认知结构体现了人类大脑网络的特征，即网络拓扑结构[9]。所有这些网络自身的约束限制以及彼此相互影响产生的动态过程使得语言成为我们今天看到的样子。

一种可靠的语言网络构造方法是语言网络研究的第一步。在某个层面上，网络假设可以被简单地看作是一种展示语言数据的标记方法。网络是节点、边的集合，构建一个网络首先要确定这两个要素[10]。迄今可见诸多对语言网络的构造，集中在以字、词为单位的语言同现网络[11-12]、句法网络[3][13]、语义网络[4][14-16]的不同层面，这些网络的构建大都受语

* **收稿日期：** **定稿日期：**

基金项目：国家社会科学基金重大项目——现代汉语计量语言学研究（NO.11&ZD188）；国家社会科学基金青年项目——基于同一文本的句法网络语义网络关系研究（NO.14CYY046）。

言资源的形式所限表现出些许差异，但已能基本窥见语言网络类似于其他自然和社会网络的统计规律（小世界、无标度）。但是如何更好地结合语言特点和语言学的研究成果，采用更可靠的方法对语言单位各层面分析，是语言网络研究者需要深入思考的问题。本文收集的几类以词为单位的语言网络构造方法,基于同一文本构建不同类型的语言网络，并试图从网络全局参数和网络局部节点特征两个角度来阐释不同层级语言网络的差异。

2 同现网络

语言同现网是语言工程领域研究者较为熟悉的网络构造方法。这种方法基于分词操作，不需要对语言单位（词）进行深入的结构分析。研究者通常先建立模型，通过模型确定词关系矩阵继而建立网络。

同现网构造方法之一是 n 阶 Markov 同现。如果在一个句子中，两个词之间在 n 阶 Markov 链的条件下存在同现关系，则认为网络中相应的两个节点之间存在一个连接。对语料库中的所有句子进行上述处理，便可构造出词同现网络。语言工程的实践表明，n 阶 Markov 链中的 n 取 2 比较合适，因为句子中两个词的邻接同现是最常见的。虽然也存在一些间隔大于 1 的相关词对，但如果在模型中考虑此种远距离关联，则会引入大量的无关词对，降低词同现网络对真实情况反映的准确性。采取这个策略，一方面可较充分地反映词与词之间的上下文制约关系，另一方面，又可使模型的复杂性得到较好的控制。

按照上述方法，我们使用两个句子的文本“人体是由数以亿计的微小而有生命的细胞构成的 这些细胞构成各个不同的组织 器官 保证了人体的正常工作 ”构建 2 阶马尔科夫链同现网络。在该同现网中节点为词，按照次序建立前后连接（箭头表示词连接方向），网络中标点符号被删除，保留句子的根节点标记 ROOT[1]，形成 23 个节点的网络如下：

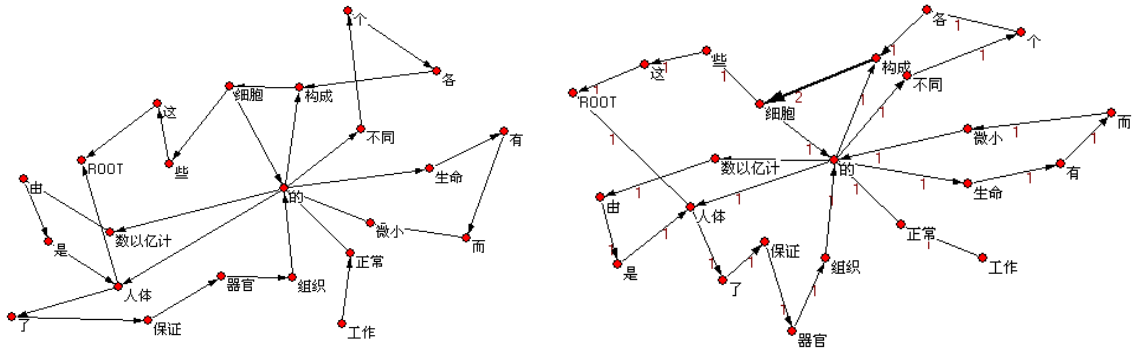

图 1　23N 有向同现网（左）和加权（频次）同现网（右）

通过 PAJEK[2]“移除多边（Net-Transform-Remove-Multiple lines）”操作，节点间在文本中对应连接的频次可以在边值中显现，文本中“细胞–构成”两词在上下文中出现，在网络中连接着两个节点的边值为 2。这样其实是构建了一个基于上下文同现的加权同现网络。

n 阶马尔科夫链构建同现网络是语言网络研究经常采用的方式，因为它的理论模型相对成熟，操作相对便捷。下面介绍的“词相似性同现网”的基本构造思想也是上下文同现，但采用了相对复杂的模型。

词相似性同现模型是 Görnerup 和 Karlgren[17]从认识语言普遍性和特殊性（它们影响分布模型行为）考虑，建立的词相似性决定的网络模型。在这个模型中，节点是词，词如果在相似上下文中同现则被连接。词相似性模型假定每个词都出现在一定的上下文概率分布之中，$P_i=\{P_i[w_p,w_i,\ w_s|w_i]\},\ w_p,w_s\in W$。在操作中，估计 P_i 通过测量 w_i 的上下文同现，再标准化这个数值。如果两个词有相似的上下文则有相似的功能。量化两个词的区别，定义 d_{ij}

[1] ROOT 为非词标记，不应存在于词节点的网络中，但考虑到其标记句子根节点的作用，且复杂网络不因单个节点的结构功能产生巨大的变化，故保留。

[2] PAJEK 社会网络分析工具。

（$0 \le d_{ij} \le 2$），通过变化相应的上下文分布来调整变化距离。词的集合和它们的相似性很容易表示为一个有权重的无向网络。节点为词，通过上下文相似度来连接。连接强度依赖于词相似度。边权通过 $w_{ij}=2-d_{ij}$ 测量。研究者测量了 11 种语言中 3000 常用词排名前 19 位的词，用它们构建词网。所有的词网有明显的社团结构，节点有组织，组织内部有高密度的边连接。社团结构的强度可以由下测量：由网络中给定的边权片段组成，这些边权来自于网络，网络中边连接相同的社团。对 11 种语言的词相似性测量发现每种语言的词相似网络都是模块化的，不同语言的模块化程度不同，芬兰语相比其他语言词间连接较弱，希腊语模块化程度明显。

不难看出无论是 n 阶马尔科夫链还是词相似性模型的同现，都是通过构造词出现的上下文环境来判断词的功能分布。但是有限元数的上下文同现难以准确反映前后成分间的规律。Liu(2008a)认为，语言同现网络的构造有其信息论的价值，但从语言学角度来分析缺乏可靠性。因为在语法上相关的成分在语序上并不一定相邻，反之，语序上相邻的成分并不一定存在语法相关性。举一个简单的例子“an interesting book”，如果在邻接的不定冠词“an”和形容词“interesting”间产生同现的连接关系可能很难找到句法理论的支持，这说明上下文同现的分析可能存在单靠词分布判断词功能的缺陷。这要求我们充分考虑句法理论在语言结构分析中的必要性。句法理论是人类（语言学家）长期的、经验的关于语言规律的总结，甚至有生物语言学的研究者主张句法是人类语言进化的结果[18]。在语言分析的时候充分考虑语言理论的研究成果是必要的，而我们目前要做的是用数学的方法和客观的数据去验证这些规律的可靠性、充分性，并通过新的大规模的数据和方法继续探索语言的规律。因此我们在构造语言网络时，有必要进入到基于句法的语言分析层面。

3 句法网络

句法网络指基于语言学（句法）理论的网络。刘海涛[16]建议构建基于语言学理论的网络，虽然从信息论角度同现网络有其价值，但是构建句法网络对于分析人类语言特征更为有益。而相比于其他句法理论，依存语法是一种“网络友好”的语言学理论[3][13][20-21]。

就句法分析而言，短语结构和依存关系是两种主要的分析手段。短语结构注重的是研究组成句子各成分之间部分与整体的关系，而依存分析关注的是构成句子各个成分之间的关系。虽然就什么是依存分析和依存语法[4][10][19][22-23]，学者们仍有不同的看法，但一般认为构成依存分析基础的是依存关系。依存关系具有这样一些主要属性：

1. 语言单位间的二元关系。这种关系在两个词间形成，也可以抽象为两个词类的间的关系；

2. 依存关系是一种有向关系或非对称关系，两个词（类）中有一个为支配词（类）。图中箭头表示这种有向性。

3. 依存关系是有标记的，即人们应该区分一种语言里的各种不同的依存关系，并且将它们显式标识出来。

依存句法理论的这些属性决定了它是一种网络友好的理论。依存句法中的词对应网络中的节点属性，关系对应边，关系类型对应边属性，这样我们就可把依存分析转化为网络。

对文本进行依存句法分析就是建立以词为单位的词间关系。对句子“人体是由数以亿计的微小而有生命的细胞构成的 这些细胞构成各个不同的组织 器官 保证了人体的正常工作”进行依存分析得到图2。

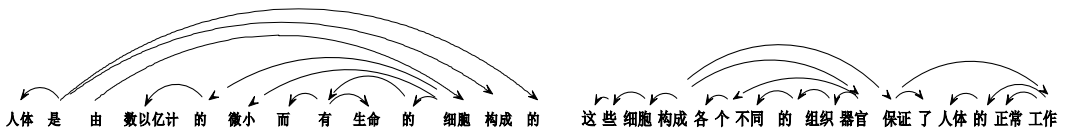

图2 线性文本间的依存句法分析

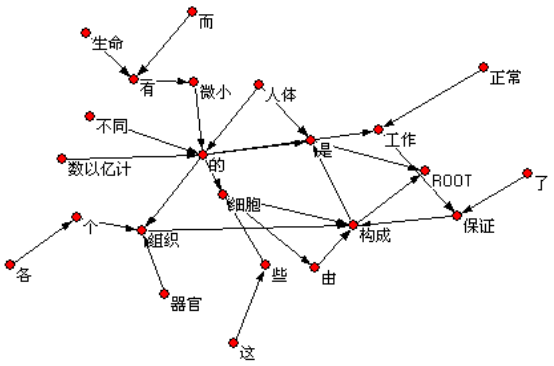

图 3　23 个节点的依存句法网络

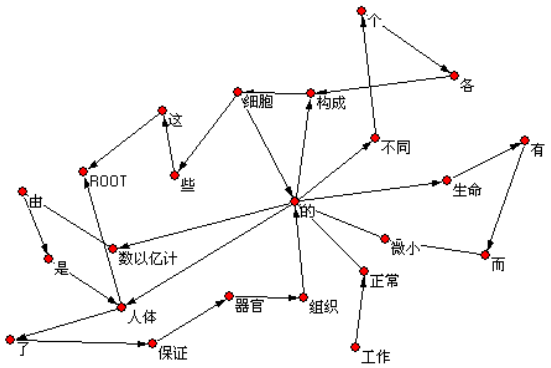

图 4　23 个节点有向词同现网

通过词间依存句法建立关系的线性文本可以容易地转化为相应语句的句法网络，如图3。这样做的优势在于：一方面，在这样的句法网络中，对文本的分析跨越了以往句法理论受限于句内障碍。另一方面，基于人脑神经网络拓扑结构的事实，如果假设文本中保持了人类的绝大多数知识，文本信息也应该存储在人脑的网状的结构中，那么，网络分析的方法实现了线性文本到人类语言存储环境（大脑）的模拟。当然，文本的网状结构并不等于人脑的神经网络结构，但是我们有理由相信文本的网状结构和人脑中知识表征、储存、学习的网状结构存在一定的联系。

为了比较同现网络和句法网络的差异，我们利用 2 阶马尔科夫链模型和依存句法理论分析构造例句的同现网（图 4）和句法网（图 3），并对两个网络的基本参数进行比较，见表 1。

表 1　23 节点同现网、句法网主要参数比较

	N	*E*	*L*	*D*	k_{in}/k_{out}	*density*	CC_1	*centralization*
同现网	23	29	4.7487	12	1.2608/1.2608	0.05482	0.0150	0.1634
句法网	23	29	2.5193	5	1.2608/1.2608	0.05482	0.0797	0.1385

注：　*N*-节点数；*E*-边数；*D*-直径；k_{in}/k_{out} t–节点入度、出度；*density*-密度；CC_1-只有1个邻居节点的聚集度；*centralization*-网络中心度

表 2　同现网和句法网标准化节点度排序

文本词频排序			同现网节点度排序				句法网节点度排序			
排序	节点号	词形	排序	节点号	词形	标准化	排序	节点号	词形	标准化
1	5	的	1	5	的	1	1	5	的	1
2	2	人体	2	1	人体	0.375	2	11	构成	0.625
3	2	构成	3	11	构成	0.375	3	10	细胞	0.5
4	2	细胞	4	10	细胞	0.375	4	2	是	0.375
5	1	不同	5	20	了	0.125	5	17	组织	0.375
6	1	个	6	19	保证	0.125	6	22	工作	0.25
7	1	了	7	18	器官	0.125	7	19	保证	0.25
8	1	些	8	17	组织	0.125	8	8	有	0.25
9	1	保证	9	16	不同	0.125	9	3	由	0.125
10	1	各	10	15	个	0.125	10	15	个	0.125
					mean:	0.1902				0.1902
					Median:	0.125				0.125
					Standard deviation:	0.1989				0.2496

在 23 个节点的有向网络中，可观察到两个网络的平均路径长度、密度、节点度相当，

而句法网路的直径 5 显著小于同现网络直径 12。虽然同现网和句法网的节点平均度整体没有差异，但是节点度分布存在明显不同。这表明句法网重新分配了词在网络中的功能。两个网络中“的”节点度排在首位。节点“的”是构建网络所用文本中最高频词，同时在句法分析中起着连接形容词和名词的重要句法作用，这是“的”在节点度分布中排在首位的两方面因素。值得注意，节点“是”在构建网络的文本中只出现一次，这影响了它在同现网节点度分布中的排序，但是在句法网络的度分布中节点“是”占据前列，这表明经句法分析构造的网络侧重反映词的语法功能价值。

在我们构建 23 节点的同现网络和句法网络度分布中，两个网络具有相同标准化度分布均值 0.1902，但是句法网络的标准差3略大于同现网络，这反映了句法网络度分布离散性较高，度分布越离散网络的层级性和异质性越高。考虑到目前网络的规模， 同现网和句法网的更显著差异可能还需要更大规模节点的网络数据支持。但是我们已经发现，这两类节点相同、组织方式不同的微型网络存在基本参数上的差别。

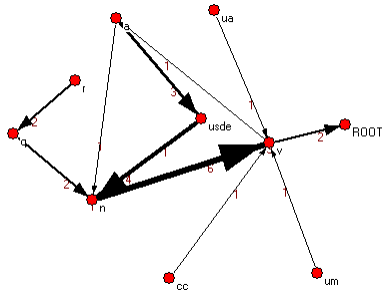

图 5 23 个节点词类网络参照

词类是句法理论研究的一项主要内容，汉语词类问题在汉语语法分析中产生的影响一直备受争议。复杂网络注重整体的特质，使得它非常适宜于研究某些词（类）对语言系统的影响。我们对文本“人体是由数以亿计的微小而有生命的细胞构成的 这些细胞构成各个不同的组织 器官 保证了人体的正常工作”进行依存分析构建了词类的关系网络，如图 5。网络中包含 10 个词类节点（和预先设定的词类分析标准有关），连接词类节点的有向边反映依存句法理论中词类间的相互支配关系（箭头所指方向关系为“从属于”），边的粗细（依赖边值）反映文本中相应类型词类间关系的出现的频次。在这样的网络中，我们能够比较直观的看到语言中哪些词类在文本比较活跃，哪些词类间存在依存关系。

这一方面最值得研究问题是汉语虚词在汉语句法体系中的作用[24]。一般认为，由于汉语的实词没有形态变化，虚词便成了汉语的主要句法手段之一。如果虚词是汉语的主要句法手段，那么从汉语句法网络中将虚词移走，可能会导致汉语句法网络的统计特征发生重大的变化。陈芯莹、刘海涛[25]以概率配价模式理论[4]为基础，利用复杂网络分析技术，研究和分析了汉语句法网络中虚词的网络结构特点。他们的研究发现：1. “的”是汉语句法网络的全局中心节点。它的被支配能力是网络中最强的，同时它还具备很强的支配能力。而且，“的”的这些网络特性受语体影响较小。从网络中剔除“的”节点，会造成句法网络的平均度下降、平均路径长度增加、直径增加、密度降低并导致孤立节点的产生；2. “了”是网络中的局部中心节点，不是全局中心节点。它具有较强的被支配能力但不具备支配能力。删除“了”会造成网络的平均度下降，但其对网络的影响比“的”要小；平均路径长度增加、直径增加、密度降低，其影响均大于“的”；不会使网络产生孤立节点；3. 介词“在”是接近

[3] 标准差（Standard Deviation）是各数据偏离平均数的距离的平均数。标准差能反映一个数据集的离散程度 。简单来说，标准差是一组数据平均值分散程度的一种度量。一个较大的标准差，代表大部分数值和其平均值之间差异较大；一个较小的标准差，代表这些数值较接近平均值。通常，标准差越高，表示实验数据越离散，也就是说越不精确。反之，标准差越低, 代表实验的数据越精确。

[4] 概率配价模式理论详见：刘海涛，依存语法的理论与实践，北京：科学出版社，2010:106-111.

网络的全局中心节点。但它的支配能力与被支配能力受语体影响较大，在书面语体中的被支配能力强于在口语体中的被支配能力。剔除“在”后，网络的平均度下降，但其影响比“的”要小；平均路径长度增加、直径增加、密度降低，其影响均大于“的”与“了”相当；会使网络产生孤立节点。

汉语依存句法网的全局特征和局部特征的研究从复杂网络和语言理论两个角度加深了我们对语言网络的认识，也促使研究者进一步探索语义网络的面貌。

4 语义网络

什么是语义网络?

与字、词、句法等表层语言网络不同，语义网络是一种深层语言网络。语义网络又可以分为两种，一种是通过真实文本进行语义角色或论元结构分析所得到的语义网络，这种网络可以称之为动态语义网络。动态语义网络有助于研究与交际过程相关的各种语义问题，有利于研究更好的语义处理策略与系统。Liu[4]通过对真实文本进行语义角色标注，构造并研究了汉语的动态语义网络。这是一种节点为实词，连接为语义或论元关系的网络。另一种是根据词典等语言资源构造的语义网络，这种语义网络是一种静态语义网络，它所反映的是人类存储知识的方式与结构。在这样的网络中，节点一般为概念（或实词），节点之间的关系可以是上下位、部分与整体、同义、反义等语义关系[26]。静态语义网络对于义类及概念词典的研究，对于知识库的开发都有用处。图 6 左是一个静态语义网络的示意图。其中空心箭头表示两词之间在语义上属于上下位关系，如“花-百合花”说明百合花是花的一种；而实心箭头表示两词之间在语义上属于部分-整体关系，如“花萼-花”说明花萼是花的一部分。在这样的网络中，节点一般为概念（或实词），节点之间的关系可以是上下位、部分与整体、同义、反义等语义关系。图 6 右是句子是小百科中关于“花”的定义“花是被子植物繁衍后代的生殖器官 一朵完整的花包括了六个基本的部分 即花梗 花托 花萼 花冠 雄蕊群和雌蕊群”中实词的动态语义网络。在这类网络中我们注意到节点不再是静态语义网络描述的同类词相关的概念网络，而是包含了多种实词类的动态网络。

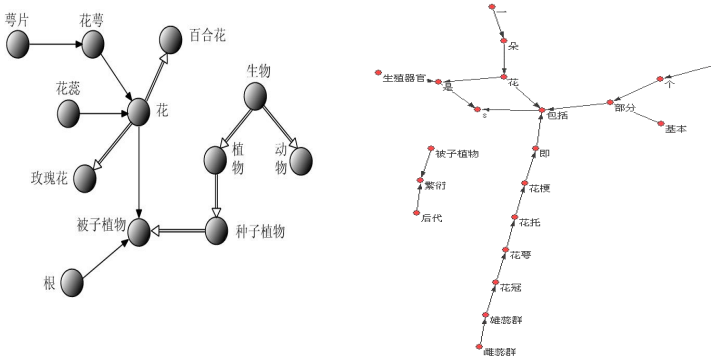

图 6 静态语义网络示意图（左）动态语义网络示意图（右）

相比静态语义网络对于义类、概念词典、知识库开发研究的作用。动态语义网络注重人在实际语言运用中对概念从语义到句法的整合和实现过程。而这个过程是认知科学、心理学、语言学共同关注的焦点。如果我们认可神经网络是人类思维的生物基础，那么就可以说，静态、动态语义网络的相互协作完成了人类思维到语言功能的实现。

通常汉语语义分析被认为是针对实词的分析。同样，在语义网络中不对只有句法功能无实义的虚词进行分析。这就涉及到汉语实词、虚词分类的问题。而虚实分类又会触及到棘手的汉语词类问题。陆俭明[27]在《现代汉语语法研究教程》中就“汉语词类问题是个老大难的问题”进行了详细论述，自中国第一部汉语语法专著《马氏文通》[5]至今已有 11 个关于汉语词类较为完整的分类体系（马氏文通、黎锦熙、吕叔湘、王力、语法讲话、中学体系、胡

[5] 参看《马氏文通》（马建忠，1989）北京：商务印书馆，2007 版.

裕树、黄廖本、朱德熙、北大本、张斌），这 11 个分类体系的看法有的部分一致、有的部分涉及词类细化、有的完全相反，对汉语虚词、实词划分也是在各自词类分析的基础上有自成一体的判断。

考虑到汉语语义网络构建过程中必然要参考汉语语法研究已有的成就，但是不宜过甚陷入学术之争。简单说，我们基本采用《中学教学语法系统提要》[6]的分类制定适用于汉语信息处理的词类标注体系，并采用其对汉语实词、虚词的分类为参考进行语义网络的提取。原因有二：一方面，中学体系影响教大，目前出版的标注词类的词典大多沿用这个体系，辞书可以为具体的语料分析操作提供详尽的有效参考；另一方面，中学体系经历长期的教学实践，较大程度决定目前国民语言文字使用的实际水平，而我们实验的语料是来源于日常使用的真实语料，采用这个系统对语料进行再分析符合构建汉语网络考察人脑对语言认知原始状态的预期。

从这两个因素考虑，我们制定汉语 12 大词类（名词、数词、量词、形容词、动词、副词、代词、介词、连词、助词、叹词、拟声词）和部分大类细分小类的标注方案[28]，认为汉语虚词是包含介词、连词、助词、副词、拟声词的类，需要明确的是，在此基础上的语义标注中“副词”类存在较大问题：黄伯荣、廖序东[29]认为副词是虚词，邵敬敏[30]认为副词兼具实词和虚词，胡裕树[31]认为副词是实词。从副词细分来看，《现代汉语副词分类词典》[32]有十小类的分法可供参考：时间副词、程度副词、限度副词（顶多、起码、大约、恰好、到处）、情态副词、语气副词（倒、到底、究竟、难道）、判断副词（的确、势必、偶尔、或许、不）、)频次副词、关联副词、目的副词、类比副词。其中否定副词“不”如果作为虚词在语义分析中提出会影响语义正确表达，在实际语义分析中我们较多遇到“不”的问题，故决定副词“不”在语义分析时保留。

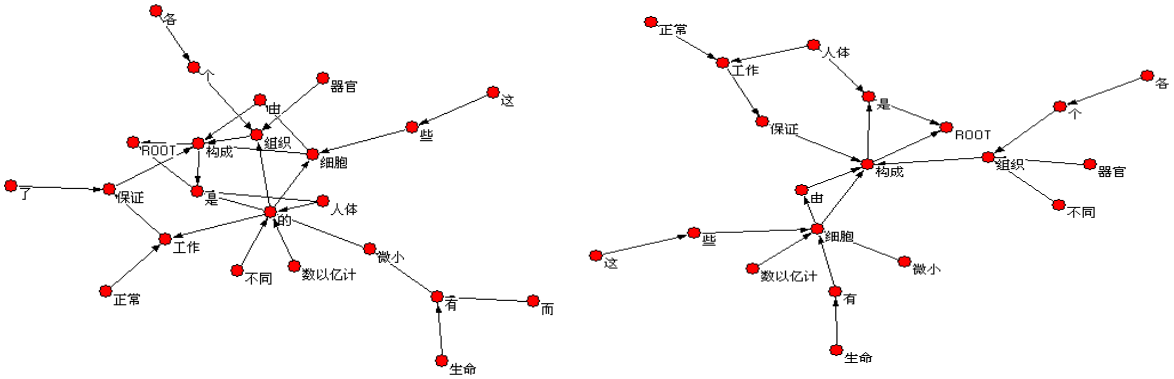

图 7　23 节点句法网（左）和 20 节点语义网（右）

利用文本“人体是由数以亿计的微小而有生命的细胞构成的 这些细胞构成各个不同的组织 器官 保证了人体的正常工作”构建的实词语义网包含 19 个实词节点，见图 7。与相应文本的句法网络相比去掉了虚词节点“的、了、而”。微型实词网络的平均路径长度、网络直径、节点入度、出度小于句法网络相应参数。平均路径长度是网络中任意两个节点之间的最短路径长度均值，它聚合了网络所有成对节点，是网络全局性指标。从网络的平均路径长度和直径来看，实词网络的密度略高，这可能和网络节点数缩小相关。但是同文本句法网络和语义网络对比同现网络，前两者具有明显小的平均路径和网络直径，如表 3 所示，这反映出从语言理论角度构建的语言网络可能具有更显著的复杂网络全局特征。语义网络的聚集系数高于同文本的同现网络，但远落后于同文本句法网络。聚集系数描述节点的相邻节点互为邻居的程度，它是反映网络中三角关系的聚集倾向和集群形态的局部特征指标。同比下句法网络具有较高的聚集系数，反映出句法网络节点间具有更为紧密的联系，去除了虚词的语

[6] 中学教学语法系统提要（人民教育出版社中学语文室，1984）根据 1981 年 7 月在哈尔滨举行的“全国语法和语法教学讨论会”上确定的原则起草。

义网络，聚集系数降低，说明虚词在连通语言网络节点局部关系上起到一定作用，这一点有待扩大网络规模后的进一步验证。

表 3 三类网络基本参数比较

	N	*E*	*L*	*D*	k_{in}/k_{out}	*density*	CC_1	*centralization*
同现网	23	29	4.7487	12	1.2608/1.2608	0.0548	0.015097	0.1634
句法网	23	29	2.5193	5	1.2608/1.2608	0.0548	0.079762	0.1385
语义网	20	22	2.2151	4	1.1/1.1	0.0550	0.0258333	0.1111

表4 句法网和语义网的中心节点标准化排序（前三位）

排序	节点	词形	标准化	排序	节点	词形	标准化
1	5	的	0.3261	1	9	构成	0.2333
2	11	构成	0.1988	2	8	细胞	0.15
3	8	有	0.055	3	19	工作	0.0566

在复杂网络尤其是社会网络分析中，网络中心（*centrality*）描述单个节点在网络中的位置，网络的中心性（*centralization*）定义整个网络的性质。如果网络的中心节点和外围节点有较为明显的界限就表示这个网络有较高的中心度。在中心度高的网络中，信息更容易传递。社会网络中，一个行动者（节点）可以通过多种途径之一占据网络中心位置：与许多其他行动者相连接（度中心性）；能接触到网络中许多其他行动者（接近中心性）；把彼此之间没有直接联系的行动者连接起来（中介中心性）；与居于网络中新位置的行动者有连接关系（特征向量中心性）。由此可见，在信息高效传递的网络中，中心节点是必不可少的。那么，语言网络作为一种包含丰富信息的网络，它的中心节点会不会因为不同的网络构造方法产生差异呢？这种差异会不会进一步导致网络局部特征和全局特征的差异呢？这些问题都有待进一步探索。在这，我们先利用 PAJEK 提取了 2 个句子文本的句法网、语义网的中心节点（Net-Vector-Centers），得到如表 4 所示排序。

在 22 个词的句法网中，“的、构成、有”具有较高的网络中心位置，其中助词“的”优势非常明显。而在去除虚词后剩余 19 个实词节点的语义网络中，中心节点发生了明显变化，句法网的中心节点“的”在实词语义网中被剔除，名词节点“细胞”在句法网中原本不具中心性，却成为了语义网络的中心节点。部分虚词和名词类中心节点的变化是句法网、语义网最显著的差异。通过网络中心节点与文本中词频的比较，我们还发现：“细胞”在文本中并非高频词，节点“细胞”能够在句法网、语义网占据网络中心位置，更多地说明名词类节点在语义网络中的重要作用。

5 结语

运用相同文本不同方法构造的小型语言网络，在网络的基本参数和网络中心节点上表现出较大差异。考虑到复杂网络技术是大规模节点计算的方法，2 个句子文本构造网络的参数测量只能算是构造语言网络的初探。小规模语言网络构造的目的是比较同现、句法、语义网络的异同，强调语言多层系统、语言学理论与复杂网络方法的联系，这是结合网络科学探究语言网络迈出的第一步。本研究还将在现有理论基础上进一步扩大语料规模以增加统计数据的有效性，观察不同规模、不同层级语言网络之间的差异，以检验网络模型应用于语言分析的可靠程度。

参考文献

[1] Briscoe, E.J. Language as a Complex Adaptive System: Coevolution of Language and of the Language Acquisition Device[C]. In: H. van Halteren et al., editor, Proceedings of Eighth Computational Linguistics in the Netherlands Conference, 1998.

[2] Steels, L. Language as a Complex Adaptive System[C]. In: Schoenauer, M., editor. Proceedings of PPSN VI, Lecture Notes in Computer Science. Berlin:. Springer-Verlag, 2000:

17-26.
[3] Liu, H. The complexity of Chinese dependency syntactic networks[J]. Physica A., 2008a, 387: 3048-3058.
[4] Liu, H. Statistical Properties of Chinese Semantic Networks[J]. Chinese Science Bulletin. 2009, 54(16): 2781-2785.
[5] Liu, H. Linguistic Complex Networks: A new approach to language exploration[J]. Die Grundlagenstudien aus Kybernetik und Geisteswissenschaft (grkg/Humankybernetik) 2011; 52(4): 151-170.
[6] Cong, J., Liu, H. Approaching human language with complex networks[J]. Physics of Life Reviews 2014, this issue.
[7] Liu H, Cong J. Empirical characterization of modern Chinese as a multi-level system from the complex network approach[J]. J Chin Linguist 2014;42:1-38.
[8] Pickering, M.J. and Garrod, S. Toward a mechanistic psychology of dialogue[J]. Behav. Brain Sci., 2004, 27: 169-226.
[9] Eguiluz, V., Cecchi, G., Chialvo, D. R., Baliki, M. and Apkarian, A. V. Scale-free brain functional networks[J]. Phys. Rev. Lett. 2005, 92: 018102.
[10] Hudson, R. Language Networks: The New Word Grammar[M]. Oxford: Oxford University Press, 2007.
[11] Ferrer i Cancho, R. and Solé, R.V. The Small-World of Human Language[J]. Proc. R. Soc. Lond. Series B, 2001, 268: 2261-2266.
[12] 刘知远, 孙茂松. 汉语词同现网络的小世界效应和无标度特性[J]. 中文信息学报, 2007, 21 (6): 52-58.
[13] Ferrer i Cancho, R., Solé, R.V., Köhler, R. Patterns in syntactic dependency networks[J]. Physical Review E, 2004, 69: 051915.
[14] Sigman, M. and Cecchi, G.A. Global organization of the Wordnet lexicon[M]. Procs. Natl. Acad. Sci. USA, 2002, 99(3): 1742-1747.
[15] Steyvers, M. and Tenenbaum, J.B. The large-scale structure of semantic networks: statistical analyses and a model of semantic growth[J]. Cognitive Science, 2005, 29(1): 41-78.
[16] Holanda, A.J., Torres Pisa, I., Kinouchi, O., Souto Martinez, A. and Seron Ruiz, E. Thesaurus as a complex network[J]. Physica A, 2004, 344: 530-536.
[17]Görnerup, O., Karlgren, J. Cross-lingual comparison between distributionally determined word similarity networks[C]. Proceedings of the 2010 Workshop on Graph-based Methods for Natural Language Processing, ACL 2010. Uppsala, Sweden, 2010: 48-54.
[18] Bickerton, D. (EDT), Szathmary, E. (EDT). Biological Foundations and Origin of Syntax (Strüngmann Forum Reports)[M]. The MIT Press, 2009.
[19] 刘海涛. 汉语句法网络的复杂性研究[J]. 复杂系统与复杂性科学, 2007b, 4(4): 38-44.
[20] Čech, R., Mačutek, J. Word form and lemma syntactic dependency networks in Czech: a comparative study[J]. Glottometrics, 2009, 19: 85-98.
[21] Ferrer i Cancho, R. The structure of syntactic dependency networks: insights from recent advances in network theory[C]. In: Altmann, G.., Levickij, V., Perebyinis, V. (eds.). The problems of quantitative linguistics, Chernivtsi: Ruta, 2005: 60-75.
[22] Tesnière, L. Eléments de la syntaxe structurale[M]. Paris: Klincksieck, 1959.
[23] 刘海涛. 泰尼埃的结构句法理论[J]. 北华大学学报(社会科学版), 2007a, 8(5): 68-77.

[24] 刘海涛. 语言网络：隐喻，还是利器?[J]. 浙江大学学报（人文社会科学版）, 2011, 41(2): 160-179.
[25] 陈芯莹, 刘海涛. 汉语句法网络的中心节点研究[J]. 科学通报,2011, 56(10): 735-740.
[26] Solé, R., Corominas-Murtra, B., Valverde, S. and Steels, L. Language Networks: Their Struc ture, Function and Evolution[R]. Santa Fe Institute Working Paper, 2005. (05-12-042).
[27] 陆俭明. 现代汉语语法研究教程[M]. 北京：北京大学出版社，2004.
[28] Liu, H., Huang, W. A Chinese Dependency Syntax for Treebanking[C]. Proceedings of the 20th Pacific Asia Conference on Language, Information and Computation: 126-133. Beijing: Tsinghua University Press, 2006.
[29]黄伯荣, 廖序东. 现代汉语[M]. 北京：高等教育出版社，1991.
[30]邵敬敏. 汉语语法专题研究[M]. 北京：北京大学出版社，2009.
[31]胡裕树. 现代汉语（重订版）[M]. 上海：上海教育出版社，1995
[32]姜汇川. 现代汉语副词分类实用词典[M]. 北京：对外贸易教育出版社. 1989.

作者简介：作者一赵怿怡（1982——），女，博士，助理教授，主要研究领域为 应用语言学，语言复杂网络。Email: zhaoyiyi@xmu.edu.cn；作者二刘海涛（1962——），通讯作者，男，博士，教授，主要研究领域为人类语言的结构模式与演化规律，语言复杂网络。 Email: lhtzju@gmail.com。