

北语学人书系

第一辑

张普

应用语言学论文集

张普 著



北京语言大学出版社
BEIJING LANGUAGE AND CULTURE
UNIVERSITY PRESS

图书在版编目 (C I P) 数据

张普应用语言学论文集 / 张普著. -- 北京 :
北京语言大学出版社, 2012.8
(北语学人书系. 第1辑)
ISBN 978-7-5619-3353-4

I. ①张… II. ①张… III. ①应用语言学—文集
IV. ①H08-53

中国版本图书馆CIP数据核字(2012)第194188号

书 名: 张普应用语言学论文集
责任印制: 姜正周

出版发行: **北京语言大学出版社**
社 址: 北京市海淀区学院路15号
邮政编码: 100083
网 址: www.blcup.com
电 话: 发行部 82303650/3591/3648
编辑部 82301016
读者服务部 82303653/3908
网上订购电话 82303668
客户服务信箱 service@blcup.com
印 刷: 保定市中国画美凯印刷有限公司
经 销: 全国新华书店

版 次: 2012年8月第1版 2012年8月第1次印刷
开 本: 787毫米×1092毫米 1/16 印张: 23.25
字 数: 371 千字
书 号: ISBN 978-7-5619-3353-4 / H·12120
定 价: 58.00元

凡有印装质量问题, 本社负责调换。电话: 82303590

出版说明

北京语言大学是一所颇具特色的学校。在这里，聚集了数百名语言教学和研究人员，语言学研究队伍极为庞大。近年来，随着中国语言文学和外国语言文学两个一级学科博士点的建立，中、外语言文学已然成为北京语言大学的两大支柱学科。依托这两大学科，一批学科带头人和学术骨干脱颖而出，其中有的已成为本专业领域的领军人物。在汉语国际教育、汉语研究、外语研究、语言信息处理、中国文学研究、比较文学研究等领域，北语学人已成为一支不可或缺和不可忽视的力量。

倏忽之间，北语建校已经五十周年。五十年来，代有才人。然而，学校一直未能对北语学人积累下来的珍贵的学术财富进行系统的梳理。为弥补此缺憾，值此建校五十周年的特殊时刻，学校决定设立“北语学人书系”，收录北语优秀学人的优秀论文，每人自成一册，不定期陆续出版。因为时间仓促，本辑只约请了已退休的博士生导师和现任博士专业学科带头人，以便能赶在校庆期间见书，初步展示北语学人的学术风貌。今后，我们仍将继续组织征集优秀书稿，以“北语学人书系”的名义分辑出版，以体现北语学术的全面性和延续性。

为在短时间内完成这批高质量书稿的征集和编辑工作，校科研处做了大量的组织宣传工作，各位作者积极甄选论文、认真校对，北京语言大学出版社的领导高度重视，编辑们付出了大量辛勤的劳动，最终使第一辑书系得以如期出版。这正是北语精神的具体体现，亦当记录并彰扬。

北京语言大学

2012年6月

目 录

动态语言知识更新专题

- 003 / 关于语感与流通度的思考
- 017 / 关于控制论与动态语言知识更新的思考
- 034 / 论历时中包含有共时与共时中包含有历时
- 047 / 关于动态语言知识更新与流通度问题
- 066 / 关于汉语语料库的建设与发展问题的思考
- 091 / 论国家语言资源
- 108 / 当前字、词、语量化研究的五个深化方向
- 118 / 论语言的稳态
- 133 / 现代汉语稳态字词语的最新监测成果
与数字化汉语教学
- 142 / 论语言的动态
- 156 / 词语发展变化研究与生命科学结合（上）
——三个最新视角与国家语言资源监测
- 161 / 词语发展变化研究与生命科学结合（下）
——三个最新视角与国家语言资源监测
- 170 / 《动态语言知识更新研究》前言
- 202 / 《动态语言知识更新研究》后记
- 206 / 2003 年主流报纸十大流行语

汉字应用研究专题

:: 汉字编码研究部分

- 217 / 必须警惕汉字编码的副作用
- 227 / 汉字编码键盘输入法面临的历史抉择
- 231 / 98 汉字编码键盘输入方法新动向
- 237 / 关于汉字键盘输入

:: 汉字研究与中文信息处理部分

- 242 / 第一部计算机自动编纂的……
- 246 / 共和国的中文信息处理 60 年

:: 汉字部件标准和规范研究部分

- 256 / 对《信息处理用 GB 13000.1 字符集汉字部件规范》部件称说的建议
- 259 / “现代汉语常用字部件名称规范”研制工作报告

语言信息处理孵化器专题

- 263 / 语言信息产业孵化器论证报告
- 272 / 关于建立语言信息处理孵化器的建议
- 276 / 给贾庆林书记的信
- 278 / 孵化器 2000 年 3 月工作备忘录

多媒体光盘专题

:: 莫必斯 (MOBIUS) 光盘国际大奖赛部分

- 281 / “莫必斯”多媒体光盘国际大奖赛简介
- 283 / 中国光盘：走上国际领奖台
- 285 / 关于“MOBIUS 1996”和多媒体光盘的评奖
- 290 / 第五届“莫必斯”多媒体光盘国际大奖赛中国为什么没有获奖？

- 298 / 一个大市场 看你怎么做
——多媒体光盘制作厂商应关注的几个问题

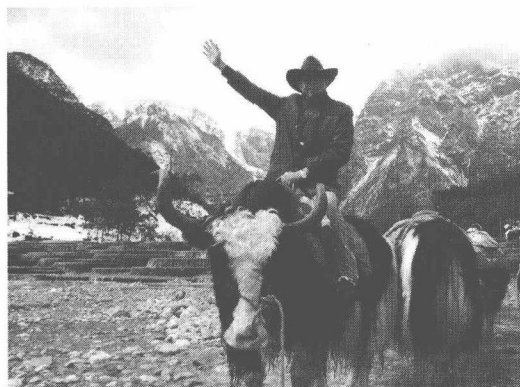
:: 多媒体光盘部分

- 302 / 数字化时代的影视珍藏
——评“芝麻开门”影视珍藏馆
- 309 / 多媒体语言教学光盘与语感能力
- 317 / 语言的多媒体性与多媒体语言知识的作用

为计算语言学专著首版所作序言专题

- 333 / 王强军《信息技术领域新术语提取研究》序
- 337 / 甘瑞瑗《国别化对外汉语教学用词表制定的研究》序
- 342 / 《行知中国》序
——行且知 行而知
- 346 / 郑泽芝《大规模真实文本汉语字母词语考察研究》序
- 351 / 刘华《词语计算与应用》序
- 354 / 杨建国《基于动态流通语料库的汉语熟语单位研究》序
——业精于勤，行成于思
- 360 / 韩秀娟《基于动态流通语料库的通用词语用字研究及字词语关系考察》序
- 364 / 后记

动态语言知识更新专题



关于语感与流通度的思考^{*}

1998年8月在应用语言学的研讨会上，我发表了一篇论文《关于大规模真实文本语料库的几点理论思考》。那篇文章对语料库的建设进行了反思，在回顾多年来语料库建设的成就和当前迫切需要解决的棘手问题的基础上，对于第三代语料库的建设，特别是面向下一个世纪的超大规模语料库的建设，从理论语言学和社会语言学的角度提出了一些想法。其中，非常重要的核心思路是提出了语言的“流通度”的概念，以及流通度的内涵和基本的参数计算公式。

本文就流通度与语感的问题进一步作一些探讨。

一、关于语感

首先，我们需要弄清楚：什么是语感？

迄今为止，语感是语言学界使用频度较高但是又琢磨不透、难于界定的一个术语。

^{*} 本文获国家自然科学基金重点项目（项目编号：69433010）资助。

《现代汉语词典》、《辞海》乃至《语言与语言学词典》竟然均未收录“语感”这个词条。《汉语大词典》收录了“语感”，解释为“对语言的感觉”，并且引证了朱自清《〈文心〉序》和陈原《社会语言学》为例。《现代汉语新词典》（刘文义主编）的解释是“对语言符号的直觉、感受及理解”。

我们检索了最近 10 年来的主要语言学刊物及学报，得到论述“语感”的论文 26 篇，这还不包括此前论及“语感”的名家的著述，如朱德熙《从作文和说话的关系谈到学习语法》（1980）和吕叔湘《中学教师的语法修养》（1985）中都谈到语感；邢公畹则在 1981 年发表《论“语感”》，该文本是为宋玉柱的《现代汉语语法论集》写的序，因为主要谈的是语感，所以单独发表时改为语感的专论。

最近一二十年，不仅是语言学界，哲学、思维科学、认知科学、信息科学等领域都有人对语感、语感能力发生浓厚兴趣。在语言学界，心理语言学、社会语言学、计算语言学、应用语言学、生成语言学等领域都从不同角度对语感、语感能力进行了界定、分类、描述等研究。尤其是语言教学领域（包括对外汉语教学），对于语感的性质、语感能力的训练、语感能力和听说读写等语言能力的关系以及语感在语言（文）教学中的地位等一系列的问题都有探讨。无论围绕着语感和语感能力有多少争论，在下述几方面应该说已经有了比较清晰的想法：

1. 语感是操某一种语言的人在长期的语言实践中形成的对于这种语言的运用的正误、优劣、常殊的一种直觉（intuition）能力。
2. 语感是一种综合的语言直觉能力。它包括语音感、语法感、语义感、语用感等许多方面的敏感的直觉能力。
3. 由于个人的文化素质、生活环境、交际范围、阅读领域、活动空间等各不相同，人们的语感能力千差万别，但是操同一语言的民族总有共同的语感和语感能力，否则就无法进行交际。
4. 语感能力也是一种语言能力，广义的语言能力可以也必须包括语感能力。听说读写是最基本的语言能力，而语感能力是听说读写能力的基础和前提。语感能力是对听说读写这种语言运用（performance）起监控作用的一种语言审析能力。
5. 语感是在长期的语言实践中逐步地、自然而然地形成的，语感可以通过语

言教学实践来培养和训练。传统的语文教学中有许多培养和训练语感的宝贵经验与方法（例如“涵泳”），如“书读百遍，其义自见”、“读书破万卷，下笔如有神”、“熟读唐诗三百首，不会作诗也会吟”都有语感的道理在其中。

总之，语感是客观存在的，语感是重要的。在语言教学中，过去对听说读写等语言运用能力较为重视，而对语感这种语言监控能力研究不够。

二、关于语感的量化

既然语感是一种重要的客观存在，并且需要加强研究，那么我们能不能对语感进行量化的研究呢？能不能使语感可计算呢？

我们从计算语言学的角度对语感的研究提出了新的目标。当然，实现语感量化和可计算，其更进一步的目的是为了让电脑模拟人的语感，或者说赋予电脑以语感，使电脑获得语言监控能力，从而具备语言的自学习、自反馈功能。

语感是对语言运用的正误、优劣、常殊的一种直觉，或者说是对表述得正确不正确（即“信”）、顺畅不顺畅（即“达”）、恰当不恰当（即“雅”）的一种直观的认识能力与审析能力。语感是一种“度”、一种“分寸”，实际上就是对信达雅的程度感，对于语言运用的分寸感。因此，语感一定是可以量化的，可以计算的。

一种语言事实能不能“被理解”、“被认同”和“被传播”，通常是看其“能不能说”。而“能不能说”主要靠的是个人的“语感”，个人的语感则与对话双方的许多背景因素有关，甚至连职业、职务、年龄、性别、性格、品德、情感、阅历等也可能影响到人们的语感。因此，可以认为：原则上没有也不大可能有两个语感完全相同的人，即使是同一个人，其语感在不同时期也可能会有所变化。

当然，操同一语言的不同的人，其语感和语感能力还是有强弱之分的。所以有的人表达起来“伶牙俐齿”，有的人则“笨口拙舌”，有的人领悟起来能够“举一反三”，有的人则必须“冥思苦想”。我们把人与人之间的这种语感的差别称为“差别语感”，用C来表示。“差别感”C是一个变数，它随着不同人的语感和语感

能力强弱的变化可大可小,同一个人的“差别感”也随着影响语感的因素的变化而变化。所以,“差别感”是语感和语感能力中的动态部分,是一种动态语感。

但是,操同一语言的人必定还有共同的语感,这样才能沟通。共同的语感是最基本的语感,最常规的语感,最一般的语感。我们把这种人与人之间存在的共同的语感称为“共同语感”,用G来表示。“共同感”G是一个定数,对于所有的人都是一定的。所以,“共同感”是语感和语感能力中的稳态部分,是一种稳态语感。任何一个人的语感(我们用Y来表示)都是有稳态部分和动态部分这样两种成分的,所以,语感的公式是:

$$Y = G + C \quad (Y: \text{语感}; G: \text{共同感}; C: \text{差别感})$$

我们说G是稳态而不说是静态,因为共同感并不是静止不变,只是相对稳定。随着时代、民族、国家的发展,随着人类的进步和世界的交融,“共同感”也会发生变化,只不过无论怎样变化,无论变化大小、快慢,对于每一个人来说,这部分永远是一样的。

我们主要分析“差别感”。人与人之间的语感差别主要体现在如下一些方面:

1. 新旧差别

这主要表现在新词、新义、新用法方面。语言总是要发展变化的,新词、新义是最活跃的部分,“能指”必须适应“所指”,新事物、新思想、新发现带来的新概念常常需要用新词、新义来表达。例如:“网址”、“彩打”、“卫视”、“手机”、“邮编”、“电喷”、“尾气”、“驾校”、“拆迁”、“拥堵”、“整治”、“监管”、“关爱”、“下岗”、“转制”、“磨合”、“超市”、“展销”、“甩卖”等等。青年人思想最活跃,最容易接受新事物,所以,一些新词、新用法往往首先从青年人群开始传播。例如:“爽”、“酷”,“帅哥”、“大款”,“网虫”、“黑客”,“蹦的”、“考托”,“光盘”、“盗版”,“帅呆了”、“酷毙了”、“很中国”等等。新的语言成分总是有人最先在局部范围内开始使用,传播开并稳定后,就成了被大家接受的新词、新义、新用法;如不然,就逐渐被淘汰,个人或极少数人仍使用,就被视为生造。语感的量化和可计算,应该能够提供一个阈值,从而较为科学地确定一个新的语言成分何时算是“约定俗成”,被公众(或者说是被绝大多数人)接受了。

2. 学科差别

这主要表现在词汇中的术语和一般词汇的术语义。一些人（往往是从事某一学科的知识分子或从事某一行业的同行）对本领域的术语和术语义有较强的语感，而该学科领域或行业范围以外的人则没有这种语感。例如：“向量”、“函数”，“扇区”、“总线”，“臭氧”、“烃基”、“等离子”、“电离层”，等等。术语或术语义的使用范围有时会扩大，进入一般词汇。例如：“攻坚”、“突击”（军事），“开刀”、“瘫痪”（医学），“票友”、“反串”（戏曲），“硬件”、“软件”（计算机），“板块”、“断层”（地理），“倒计时”、“软着陆”（航天），等等。进入一般词汇后，学科差别消失，成为新的具有共同感的词汇和词汇义。语感的量化和可计算，应该可以寻找到从术语进入一般词汇的阈值。

3. 地域差别

这主要表现在方音、方言词汇、方言义、方言语法等方面。一些人长期生活在方言区，对本方言区的语感较强，不在该方言区生活的人或者不了解该种方言的人，就缺乏或不具备该方言语感。方音、方言词汇、方言义或方言语法有时也会扩大，进入普通话领域。例如：“搞”、“尴尬”、“垃圾”、“靓”、“焗油”、“硬是”、“有售”、“打工”、“炒鱿鱼”等等。进入普通话后，地域差别消失，成为新的具有共同感的词汇、意义和用法。语感的量化和可计算，应该也可以帮助我们寻找到一个语言成分从方言进入普通话的阈值，达不到这个阈值的就仍然是方言成分，例如：“搞掂”、“早晨（‘早安’义）”、“困觉”、“行街”、“有去”、“过夜（‘吃晚饭’义）”等等。

4. 文白差别

这主要表现在文言词语、文言用法在现代汉语中的遗留。一些人出于不同的交际目的或者处于不同的交际环境，常常使用某些文言词语或文言语法，不熟悉文言的人，就不会有文言的语感。例如：“兹”、“为荷”，“国是”、“丁忧”，“动之以情、晓之以理”、“是可忍，孰不可忍”、“勿谓言之不预也”，等等。平常说这叫“转文”。“白”是指口语，不用口语而用文言，这是一种以文白来表现的风格

差别。在特定的场合需要体现出文白的差别,以显示学问的高深或文笔的凝练、庄重等。怎样确定一个语言成分属文言还是白话?现代书面汉语中究竟有多少文言成分的遗留?语感的量化和可计算应该也可以帮助我们找到区分文言和白话的阈值,以确定哪些是文言成分的遗留。

5. 翻译差别

这主要表现在外来词的翻译方面。最初,不同的人或不同的地区对同一事物常常会有不同的译法。例如:“BB机”、“BP机”、“拷机”、“寻呼机”、“呼机”,“大哥大”、“移动电话”、“手提(持)电话”、“手提”、“手机”,“互联网”、“国际网”、“网际网”、“国际互联网”、“因特网”,等等。以后,也许会慢慢趋于一致,或者由有关的标准和规范来统一。例如:取“邮编”废“邮码”,取“因特网”,废其他。此外,欧化的句式也是出于翻译的影响。语感的量化和可计算,应该也可以帮助我们找到合适的阈值,来确定外来词的恰当译法。

也许语感还有其他方面的差异,暂不细论。

共同语感 G 总是稳态的语感,是操同一语言的全体人的语感。差别语感 C 总是动态语感,是操同一语言的部分人的语感,这“部分人”有三种情况:少部分人、大部分人、一半人。刚好一半人说的的时候是偶然的,少部分人、大部分人说的时候是必然的。只有当一种语言现象为大部分人或者绝大部分人所接受时,这种语言现象才可以说是比较成熟的,已经流通了。如果需要进行语言规范,除了共同感的部分进入规范外,这种已经为大部分或绝大部分人接受的语言事实也具有了进入规范的资格。一大批这种“多数人认同”的语言事实进入规范后,自然对这种事实语感不强的少数人会反对,那是很正常的。只不过我们无从证明反对的人就肯定是代表少数,所以规范常常陷入两难的境地。当然,反对者当时可能代表多数,时过境迁后又可能成为少数,语言随时都在发生变化,这也常常使我们的一些语言学家陷入尴尬的境地。例如,当初被极力禁绝的“人均”、“达标”、“彩电”、“面的”、“邮编”、“焗油”等,最终顽强地生存下来,有的甚至进入词典或成为规范。

倘若语感可以量化，可以计算，我们就能科学地找到规范化的“多数”的依据，从而以“(数)理”服人，而不仅仅是以权力或权威服人。当然，如前所说，我们进一步的目标是让电脑模拟人脑。赋予电脑语感，实际上就是给它一个“大多数”的阈值，叫电脑能科学地判定那些“成熟的”、“已经流通的”语言事实，自动地“认可”和“接受”，这就是自学习和自反馈，就是人工智能研究所追求的高智能境界。

但是，让电脑获得语感谈何容易。不说别的，单就科学地确定那个代表“大多数”的阈值，就不是一件容易的事。1998年，中国科协对公众的科学素养及其对科学技术的态度作了一次调查。其中，调查了公众对“分子”、“计算机软件”和“DNA”三个科技名词基本概念的理解，公众自报了解这三个术语的科学概念的分别为32.8%、15.8%和13.3%；对这三个术语不了解的公众分别为36.9%、54.4%和27%；而从未听说过“分子”一词的竟还有25.6%。仅依据百分比而言，了解这三个词的公众都不到“多数”，更不到“大多数”，那么，这三个词则都还属于科技术语，尚未进入通用词汇范围。（滕文渊，1998）但是，实际上会有许多人关心：此次抽样的原则、范围、方法是什么？文盲、家庭妇女、大学生、教授、科学家等不同人员的比例和权重一样吗？如果不一样又如何确定不同人员的比例和权重？自报“了解”的依据又是什么？等等。可见语感的量化和计算本身就不是一件容易的事，让电脑获得语感就更是困难。

不过，困难并不等于不可能，只是需要认真研究，选择一条可以操作的途径，来对语感进行量化和计算。

这就必须涉及流通度及流通度与语感的关系了。

三、流通度与语感

倘若我们要知道一个语言事实是否被大多数人认可和接受，即符合大多数人的语感，最科学的办法是向操这种语言的所有人进行调查，但这实际上是不可行的，因为：

1. 就算抛开港澳台和海外华人暂不论, 仅以大陆而言, 我们实际上无法让 13 亿人就一个语言事实来回答 “Yes” 和 “No” 或 “是” 和 “否”。

2. 何况要调查的语言事实并不是一个, 而是一长串、一长篇或一大本, 这不是 13 亿人每一位都能够回答、都愿意回答的。

3. 那一大串、一长篇或一大本需要调查的语言事实, 由谁提出? 怎样提出? 并不是随便谁任意提出一堆语言事实大家就可以认可进行调查的。

4. 更何况要保证这种调查的科学性, 实际上还应该不间断地进行调查, 因为新的语言成分随时都在产生和增长。我们也不可能一而再、再而三地时常拿这样的问题去打搅 13 亿人。

当然, 很容易想到的可行办法就是在 13 亿人中进行抽样问卷调查。但那必须具备以下的条件或前提:

1. 有人愿意提供足够的常年语感调查经费。

2. 有人能够并愿意接受常年的语感问卷调查。

3. 能够科学地确定一个从各个角度看都具有代表性的被试名单, 抽样的各种代表的比例和权重首先需要被认可和接受。

这些条件和前提中的任何一条实现起来都是很困难的。因此, 我们希望能够寻找测量语感的其他方法。

我们发现语感与流通度之间有十分密切的关系。

我们提出 “流通度” 概念, 希望通过测量流通度来对语感加以数学界定、加以量化, 使得 “能不能说”, 是否已经 “被理解”、“被认可”、“被传播” 变得可以通过流通度的计算进行判定。进一步还想把 “流通度” 的知识或者说 “流通度” 的获取方式教给电脑, 使电脑通过获得 “流通度” 来获取 “语感”, 或者来自动计算语感, 从而获得自学习能力。所以流通度理论不仅是在语言学方面使人的 “语感” 得以量化, 更重要的是在信息处理方面有可能使计算机真正获得语言的自学习能力, 使智能化进入一个新的发展阶段。

那么, 什么是 “流通度” 呢? 简单而通俗的解释就是: “流通度” 是一种语言事实在社会交际中的流行通用的程度。人们对一种语言现象的流行通用程度的感觉, 也就是所谓的 “语感”。流行通用程度高, 听得多, 习惯成自然了, 就感

觉能说，否则，就觉得不能说。“语感”处于似有似无或不高不低之间就“拿不准”。

如果我们能够对某种语言事实进行“流通度”的测量，得知这种语言事实在社会中流行通用的程度，这种“度”也就代表了所有操这种语言的人使用该语言成分的情况，对这种“度”的感觉就是语感。我们发现“度”的高低刚好与语感代表的人是“大部分”还是“少部分”成正比，因此，测到了“流通度”也就大体测到了语感。

那么，如何测量流通度呢？

四、关于流通度与动态语料库

在谈论如何测量“流通度”这个关键问题之前，我们还需要界定一下测量的对象。我们此前界定的“语感”是对语言运用的一种直觉，“流通度”是语言事实在社会中的流行通用的程度，而语言是有口语和书面语之分的。口语的语言成分“流通度”（或“语感”）的测量，目前的技术难度较大，本文主要论及书面语的“流通度”的测量方法。

关于口语和书面语的界定，我们接受汉字改革的先驱者之一周有光先生的提法，他认为书面语有两种：一种是文言，是以古汉语为基础经过加工的书面语，与口语的距离很大；一种是白话，是以现代汉语为基础经过加工的书面语，与口语一致。白话也经过了加工，加工体现在两方面：一方面是写共同语（普通话），不写方言；另一方面是对口语要作语法修辞规范化的加工。对于后一种书面语，他认为：“写出来是‘语体文’，说出来是‘文体语’，‘言文一致’。”（周有光，1998）

我们要测量的“书面语”的流通度，就是周有光先生说的第二种书面语，并且是写出来的“语体文”。他指的说出来的“文体语”和未加工的真实“口语”我们都暂且不论。

我们实际上是主张建立一个动态的大规模真实文本的书面语的语料库，把