

合成词加工中的词频、词素频率 及语义透明度^{1)*}

王春茂

彭聃龄

(中国科学院心理研究所, 北京, 100101)

(北京师范大学心理系, 北京, 100875)

摘 要 通过变化汉语合成词的词素累积频率的方法, 考察表面频率和语义透明度对加工速度的影响。研究发现, 当控制透明度时, 高表面频率词中存在词素累积频率效应, 即词素频率越高反应越快; 而在低表面频率词中没有。当控制表面频率、变化词素频率和透明度时, 在透明词中, 词素累积频率高的比低的加工快; 但在不透明词中, 这种反应趋势反转了: 词素频率越高反应越慢。这项研究表明词和词素单元表征于同一层次, 都从下一层次得到激活, 且在词和词素之间有连接; 在透明词中这些连接是正性的, 在不透明词中这些连接是负性的。

关键词 通达表征, 表面频率, 词素累积频率, 语义透明度。

分类号: B842.1

1 引 言

一般认为, 词语的视觉加工要经过视觉特征分析、词条通达、词义激活等几个阶段。而其中一个承上(整合特征分析的结果)、启下(连接相应的意义)的层次就是通达表征^[1]。用局部表征的观点可以把它看作一个抽象的词或词素结点的集合。词形信息经视觉特征分析后传至通达表征, 激活相应的结点; 被激活的结点再将信息传至语义层次, 激活相应的语义表征。

有关通达表征的结构主要有三种观点: 词素分解存储、整词存储和混合存储。词素分解存储的通达表征认为, 在通达表征层次中只有不可再分的词素, 而没有彼此独立的词条^[2,3]。也就是说, 词语是以词素分解形式存储在通达表征中的。这种表征方式的一个突出的优点是经济, 用少量的单元就可表征大量的词汇。然而它无法解释不透明词如何通过其词素获得该词的语音和语义的问题(binding 问题)^[4]。比如, 不透明词“马虎”的意思不能通过对词素“马”和“虎”的意义的整合而获得, 因此有必要引入整词的表征。整词存储的通达表征认为, 在通达表征中存储的都是整词, 每一个词都有独立的词条^[5,6]。单词的识别就是通达到其对应的词条。整词表征不存在 binding 问题, “马虎”的意义可通过其在通达表征中的独立的结点映射到相应的语义表征。然而它意味着表征的冗余, 被指

1) 本文初稿收到日期: 1998-06-20, 修改稿收到日期: 1998-09-24。

* 本研究得到国家自然科学基金会39470250, 397711120158项目和教育部九五规划项目的资助。实验在北京师范大学认知开放实验室完成。

责为不经济。混合的通达表征认为词素和整词都有可能是通达表征中的单元^[7-9]，它具有更大的灵活性。

本研究包括两个实验。实验一用变化词素累积频率和词的表面频率的方法，考察词的表面频率对通达表征的影响。实验二用变化词素累积频率和词的透明度的方法，考察词的语义透明度对通达表征的影响。所谓表面频率，指的是合成词作为一个整体出现的频率；词素累积频率指的是该词素在不同的词语中出现的频率的总和，在汉语中通常对应的是字频。

语义透明度评定

所谓语义透明度，指的是复合词的语义可从其所组成的各个词素的语义推知的程度，其操作性定义为整词与其词素的语义相关程度。实验选材中透明度得分的依据是刘颖用 9 点量表所评定的结果。量表中包含 1500 个双字词，被试为 200 名北京师范大学的本科生。要求他们在一个包含 150 个词的 9 点量表上对词本身与第一词素和第二词素的意义相关程度作出评定。如对“草率”，评定“草”与“草率”、“率”与“草率”的语义相关程度。每个词由 20 个被试进行评定，对 20 个评定分数进行平均作为得分。对第一和第二词素的平均得分作为整词的语义透明度得分。

2 实验研究

2.1 实验一：词素频率和表面频率的影响

词的表面频率的作用是一个非常强的现象，已为众多的实验研究所证实。词素累积频率的作用也在许多实验中得到验证。Zhang 和 Peng 等^[10]发现在汉语双字词的词汇判断作业中词素的累积频率的作用，Taft^[11]也发现了词素频率的作用，但是，还没有研究在一个实验里同时考察词的表面频率和词素累积频率的作用。然而，正是对词的表面频率对词素累积频率的作用的分析，有望揭示在通达表征中词和词素的关系。因此在本实验中，我们力图通过控制整词表面频率和词素累积频率的方法，揭示通达表征中词和词素的关系。本实验与以往的研究不同的地方还表现在对透明度的控制上。在 Zhang、Peng 以及 Taft 等人的研究中，透明度没有作为一个潜在的变量加以控制，因此难以断定其结果中的不一致究竟说明了什么。在本实验中，除了将表面频率和累积频率作为自变量加以控制以外，各个实验组的透明度也得到平衡。

2.1.1 方法

被试：北京师范大学的 21 名本科生参加了该实验。他们的视力或矫正视力正常，均为右利手。

材料和设计：实验中用了 40 个高频和 40 个低频双字词（表面频率高、低两组）。根据组成这些词的词素的频率高低，每一组材料又分成高词素频率组和低词素频率组（参见附录 1）。这 4 组材料在透明度、笔画数、部件数等因素上都得到了平衡（ $P > 0.20$ ，见表 1）。

我们没有匹配字的组词频率，因为一些实验表明组词频率与词素累积频率有很高的相关，前者的单独效应不显著。我们还人为地造了等数量的非词。为避免被试形成某种特定的策略，所有非词都在词素频率上和真词匹配。真词和非词随机混合成一个刺激呈现序列，真词或非词的连续呈现都不超过 3 次。实验数据开始收集之前另加了 4 个刺激，其

中 2 个真词, 2 个非词。在正式实验之前有一个练习, 材料为 18 个真词和 18 个非词。所有呈现的刺激中都无重复的字。

表1 实验一中材料的匹配信息(表面频率单位: 1/131万^[12]; 词素频率单位: %^[13])

词素频率/ 表面频率	表面 频率	首字 频率	尾字 频率	首字 透明度	尾字 透明度	首字 笔画数	尾字 笔画数	首字 部件数	尾字 部件数
低/低	6.80	0.0624	0.0934	4.995	5.294	9.300	8.800	2.750	2.800
低/高	42.60	0.0947	0.0810	5.882	5.589	8.316	9.158	2.789	2.737
高/低	5.65	0.8367	0.9157	4.935	5.129	8.100	8.900	2.400	2.750
高/高	40.50	0.7557	0.9630	4.951	5.380	8.350	8.350	2.600	2.650

仪器: 刺激的呈现和数据的收集都在 IBM 兼容机上由 Dmaster 软件完成。刺激呈现在 VGA 监视器的中央, 黑底白字。主试在另一个房间的另一台监视器前观察被试的反应。

实验程序: 被试坐在监视器前, 要求对所呈现的两个汉字尽快、准确地作出真假词的判断。如果是词, 就用利手(全为右利手)作出肯定的判断; 如果不是词, 就用非利手(左手)作出否定的判断。实验的进程如下: (a) 监视器中央呈现注视点; (b) 随后是 300 毫秒的间隙; (c) 呈现实验刺激(两个汉字) 500 毫秒; (d) 空屏 3000 毫秒; (e) 被试在 (c) 和 (d) 之间作出反应。

2.1.2 结果

只有真词的数据参加了分析。反应时大于 2000 毫秒和小于 150 毫秒的数据被看作是错误反应。去掉的数据占总数据的 0.24%。反应时的平均数加减两个标准差作为临界值, 在此临界值以外的数据由临界值代替。按这个标准修改了 2.38% 的数据。这样处理后, 所有正确的反应时和错误率分别按被试和项目求得平均数, 参与随后的被试分析项目分析。全部结果见表 2。

表2 不同表面频率和累积频率下的词汇判断的反应时和错误率

词素累积频率	反应时(毫秒)		错误率(%)	
	低	高	低	高
表面频率低	567.19	567.04	5.48	5.00
表面频率高	534.75	515.49	2.14	1.43

按被试和项目求得的反应时和错误率的平均数参加了以下的两个方差分析, 其中词素累积频率和表面频率作为被试内变量。一个是把被试看作随机变量的重复测量方差分析(以下简称被试分析); 另一个是把项目看作随机变量的独立组方差分析(以下简称项目分析)。

反应时的分析发现, 整词的表面频率的主效应非常显著[项目分析: $MSE = 1940.87$, $F(1, 76) = 18.73$, $P = 0.000$; 被试分析: $MSE = 37722.76$, $F(1, 20) = 64.38$, $P = 0.000$]。词素累积频率的主效应在被试分析时显著 [$MSE = 2633.46$, $F(1, 20) = 10.52$, $P = 0.004$], 项目分析不显著 [$MSE = 1940.87$, $F(1, 76) = 1.04$, $P = 0.312$]。词素频率和表面频率之间的交互作用在被试分析时显著 [$MSE = 1628.88$, $F(1, 20) = 4.11$, $P = 0.056$]; 项

目分析不显著 [$MSE = 1940.87, F(1, 76) = 1.03, P = 0.313$]。无论是被试分析还是项目分析, 低表面频率组里都没有发现显著的词素频率作用的简单效应 ($F's < 1$)。高表面频率组里高词素累计频率组的反应比低词素累计频率组快 20 毫秒, 且在被试分析时达到了显著水平 [$MSE = 4202.31, F(1, 20) = 23.15, P = 0.000$], 尽管项目分析不显著 [$MSE = 2387.67, F(1, 76) = 1.68, P = 0.199$]。

平均错误率的被试分析和项目分析都没有得到显著的交互作用 ($F's < 1$)。表面频率的主效应在被试分析和项目分析时都显著 [$MSE = 10.01, F(1, 20) = 8.62, P = 0.008$; $MSE = 1.06, F(1, 76) = 9.91, P = 0.002$]。词素频率的主效应不显著 ($F's < 1$)。

2.2 实验二: 词素频率和语义透明度的影响

实验二的主要任务是要探讨语义透明度对词语的通达表征结构的影响。如果透明词和不透明词在通达表征上有所差异, 那么这种差异有可能在词素累积频率效应上体现出来。

实验中我们严格平衡了词的表面频率和其他一系列因素, 变化词的透明度和词素频率。实验任务仍然是词汇判断, 着重考察透明度与词素频率之间的交互作用。

2.2.1 方法

被试: 北京师范大学的 17 名本科生参加了该实验。他们的视力或矫正视力正常。

材料和设计: 实验中用了 40 个语义透明的和 40 个语义不透明的双字词。根据组成这些词的词素的频率高低, 每一组材料又分成高词素频率组和低词素频率组 (参见附录 2)。这 4 组材料在表面频率、笔画数、部件数等因素上都得到了平衡 ($P > 0.10$, 见表 3)。另外人为地造了等数量的非词, 方法同实验一。

表3 实验二中材料的匹配信息 (表面频率单位: 1/131万^[12]; 词素频率单位: %^[13])

词素频率/ 透明度	表面 频率	首字 频率	尾字 频率	首字 透明度	尾字 透明度	首字 笔画数	尾字 笔画数	首字 部件数	尾字 部件数
低/低	23.25	0.0752	0.1004	4.083	3.956	8.947	8.800	2.700	2.800
低/高	20.70	0.0822	0.0734	6.839	7.005	8.600	9.900	2.850	2.900
高/低	24.30	0.6712	1.0316	3.644	3.924	8.350	8.200	2.550	2.600
高/高	23.60	0.8306	0.7744	6.219	6.748	8.050	8.850	2.450	2.650

仪器和实验程序: 同实验一。

2.2.2 结果

与实验一相同, 只有真词的数据参加了分析。反应时的调整方法也和实验一相同。由此调整的数据占 1.84%。平均的反应时和错误率见表 4。

对平均反应时和错误率作 2×2 的方差分析。反应时的词素频率的主效应在被试分析时显著 [$MSE = 2395.57, F(1, 16) = 10.74, P = 0.005$], 项目分析时不显著 [$MSE = 2073.49, F(1, 76) = 1.18, P = 0.280$]。透明度的主效应被试分析和项目分析都不显著 [$MSE = 59.55, F(1, 16) = 0.29, P = 0.598$; $MSE = 2073.49, F(1, 76) = 0.01, P = 0.922$]。

反应时的被试分析和项目分析都发现了显著的交互作用 [$MSE = 20613.31, F(1, 16) = 39.75, P = 0.000$; $MSE = 2073.49, F(1, 76) = 12.17, P = 0.001$]。简单效应分析发现在

表4 不同透明度和累积频率下的词汇判断反应时和错误率

词素累积频率	反应时(毫秒)		错误率(%)	
	低	高	低	高
不透明	516	553	2.94	4.12
透明	540	506	5.88	2.06

透明词中有显著的词素累积频率效应[被试分析: $MSE = 18531.58, F(1, 16) = 36.93, P = 0.000$; 项目分析: $MSE = 2046.82, F(1, 76) = 10.61, P = 0.002$]。在不透明词中, 这种简单效应的被试分析显著 [$MSE = 4477.30, F(1, 16) = 18.67, P = 0.001$], 项目分析接近显著 [$MSE = 2046.82, F(1, 76) = 2.92, P = 0.091$]。

我们注意到, 词素累积频率效应在透明和不透明词中的效应是不一样的。在透明词中, 高词素累积频率词的反应时要比低词素累积频率词快 37 毫秒; 在不透明词中, 高词素累积频率词要比低词素累积频率词慢 24 毫秒。

对错误率的分析发现词素频率的主效应的被试分析和项目分析都不显著 ($F's < 1$)。透明度的主效应被试分析和项目分析时也都不显著 [$MSE = 1.19, F(1, 16) = 1.27, P = 0.277$; $MSE = 1.08, F(1, 76) = 0.94, P = 0.335$]。

被试分析发现显著的交互作用 [$MSE = 4.25, F(1, 16) = 4.86, P = 0.043$]。简单效应分析表明词素累积频率效应在不透明词中是显著的 [$MSE = 2.94, F(1, 16) = 4.68, P = 0.046$], 在透明词中接近显著 [$MSE = 1.44, F(1, 16) = 3.27, P = 0.090$]。透明词中高词素累积频率词的错误率比低词素累积频率词的少 2.06%; 不透明词中高词素累积频率词的错误率比低词素累积频率词的多 2.94%。尽管项目分析的交互作用接近显著 [$MSE = 1.08, F(1, 76) = 3.36, P = 0.071$], 但没有得到显著的简单效应 ($P's > 0.10$)。

3 讨论

实验一发现词的表面频率和词素的累积频率对汉语双字词的加工时间都有影响。事实上, 尽管近 20 年许多研究者都发现了大量的词素累积频率效应 (CMFE, cumulative morpheme frequency effect), 他们中绝大多数也得到了明显的表面频率效应 (SFE, surface frequency effect)^[10,14], 但他们都没有研究表面频率与词素累积频率的关系, 因此, 我们不能武断地说通达表征的单元是词素或词。一个合理的推断是词和词素都存在于通达表征之中。

实验一中发现在高表面频率组中高词素累积频率的单词的词汇判断时间要比低词素累积频率的单词的词汇判断时间短 20 毫秒; 但是, 在低表面频率组中这种差异消失。如果把词素累积频率效应看作是单词分解表征的特点, 那么这个实验的结果就暗示高频词是分解加工的, 而低频词是整体加工的。这种解释不能令人满意。因为, 我们一般认为高频刺激趋向于整体加工, 低频刺激趋向于分解加工。当不同刺激经常地同时出现时, 倾向于形成一个独立的单元 (如 AAM 模型^[7,15,16])。

一个简单的混合的通达表征不能解释此处得到的结果。也许通达表征并不是一些单元的简单堆积, 单元之间有着复杂的连接。我们假设在高表面频率词得到的词素累积频率效应来自于词素和其组成的词之间的较强的连接, 而在低表面频率词中, 这种连接比较

弱,因而词素对其整词的作用比较小,没有出现词素累计频率效应。用连接主义的观点能很好理解这种连接的形成。一个词出现得越多,它和它的词素的连接就得到越多的强化。

这里又有一个问题:整词和其词素的连接是自下而上的还是同一层次内部的连接?如果按照 Taft 的交互激活模型^[11]和 Zhou 等的多层聚类模型^[17,18],合成词的加工要依次经过词素层和整词层,那么整词和它的词素的联系会影响到这个词本身的通达:低表面频率的词的加工要慢于高表面频率的词的加工。词素累积频率的作用也可归因于词素层次。这和实验的结果相符。但是,在解释表面频率和词素累积频率之间的交互作用时,这类模型遇到了困难。如果词素和整词分属不同的层次,那么它们的效应应该是独立的,不应该有交互作用。换句话说,词素累积频率效应不仅应该存在于高频词中,也应该存在于低频词中。这与实验的结果不相符。

另一种观点是将词和词素放在同一层次,如 AAM 模型^[7]。这类模型可以很容易地解释表面频率效应和累积频率效应,它们都是由于通达单元的不同激活阈限造成。但是,在解释高低词频的词的表征差异时,该模型认为高频词是整体表征,低频词是分解表征,这与实验的结果相矛盾。

如果在这类模型中加入词和词素间的连接,就可以很好地解释高表面频率词中出现的累积频率效应。通达表征中的词和词素单元都接受从低级的刺激特征层次传来的激活,词单元还接受来自于同一层次的词素单元的激活。高表面频率词和词素的联系较强,低表面频率词和词素的联系较弱。因此,高频词受到词素的影响较大,低频词受到词素的影响较小。表现在词汇判断任务中就是高频词中有词素累积频率效应,低频词中没有。

另外,如果我们要区分词素和整词是否在同一层次还可从另一个角度加以考虑。假如词素是加工到词的必经之路,那么词素的频率对词的通达的影响将不会受到其他因素的左右。也就是说,当词的表面频率平衡时,高词素频率词的加工必定要快于低词素频率词的加工。相反,如果词素和词在同一层次,它们之间的联系可能受到更上一层信息的影响,即语义层的信息有可能影响到通达表征的结构。这也就是我们所关心的语义透明度的问题。透明词的语义和组成它的词素的意义有较大的相关,不透明词的语义和词素的意义则关系不大。这种差异是否会影响到通达层次上词和词素的连接呢?

实验二的结果很好地回答了我们所提出的问题。很明显,语义透明度对词素累积频率效应的影响非常显著。尽管反应时和错误率中都发现了显著的词素累积频率效应,但必须注意的是,这种效应在透明和不透明词中的作用是相反的。在透明词中,词素频率高的词的加工要快于词素频率低的词;但是在不透明词中,词素频率高的词的反应反而变慢了。这显然不支持认为词素和整词属于不同层次的观点。因为若是那样,词素频率高的词的通达应该快于词素频率低的词。

词素频率高的词的反应不如词素频率低的快在其他人的研究中也类似的发现。Taft 等^[19]在一个实验中发现词素累积频率低-低组的反应时和高-高组差不多,并都要快于高-低组和低-高组。他们将这种奇怪的现象归因于实验所用的低-低组词的特点:大多数实验用词都是单词素词。在实验二中,这个问题不存在。所有的真词都是多词素词。

那么怎样解释在实验二中发现的透明词和不透明词的相反的词素累积频率效应呢?如前所述,我们认为通达表征是混合的,词素和词都在词汇通达的同一层次。在通达层次

上,词和它们的词素之间存在连接。从本实验所得到的结果中我们可以看到,透明词的词素频率越高,整词的反应越快,这表明透明词的词素对整词起促进作用;然而,不透明词的词素频率越高,整词的反应越慢,表明不透明词的词素对整词起抑制作用。也就是说,在词素单元和整词单元之间有着某种联系,且这种连接在不同透明度的词中是不同的:在透明词中是正的;在不透明词中是负的。在词语的加工过程中,词素单元和整词单元同时受到来自特征层次的激活,同时整词单元还受到来自词素单元的激活。对于透明词,来自于词素单元的激活是正性的,从而促进它的加工,因此词素频率越高,加工越快;对于不透明词,来自于词素单元的激活是负性的,从而抑制它的加工,因此词素频率越高,加工越慢。

综上所述,我们得出如下结论:心理词典中整词和词素的表征都存在;词和词素之间有着某种联系:其强度受词频的影响,词频越高,联系越强,对于低频词,词素和整词的联系很弱,对整词的加工影响不大;其性质受语义透明度影响,透明词的词素促进整词的加工,不透明词的词素抑制整词的加工。

致谢

刘颖在透明度的测量中做了大量的工作。刘颖、丁国胜、Marcus Taft等给作者提出了许多有价值的建议。在此一并致谢。

参 考 文 献

- 1 Sandra D. The morphology of the mental lexicon: Internal word structure viewed from a psycholinguistics perspective. *Language and Cognitive Processes*, 1994, 9(3): 227—269
- 2 Taft M, Forster K I. Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 1975, 14: 638—647
- 3 Taft M, Forster K I. Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, 1976, 15: 607—620
- 4 Rumelhart D E, McClelland J L, the PDP Research Group (Eds). *Parallel Distributed Processing*. Vol. 1. Cambridge, MA: MIT Press, 1986
- 5 Lukatela G, Gligoric B, Kostic A, Turvey M. Representation of inflected nouns in the internal lexicon. *Memory & Cognition*, 1980, 8: 415—423
- 6 Fowler C A, Napps S E, Feldman L. Relation among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition*, 1985, 13: 241—255
- 7 Caramazza A, Laudanna A, Romani C. Lexical access and inflectional morphology. *Cognition*, 1988, 28: 297—332
- 8 Feldman L B. Beyond orthography and phonology: Difference between inflections and derivation. *Journal of Memory and Language*, 1994, 33: 442—470
- 9 Schriefers H, Zwitserlood P, Roelofs A. The identification of morphologically complex spoken words: Continuous processing or decomposition? *Journal of Memory and Language*, 1991, 30: 26—47
- 10 Zhang B Y, Peng D L. Decomposed storage in the Chinese lexicon. In: Chen H C, Tzeng O J L (ed). *Language processing in Chinese*. Amsterdam: North-Holland, 1992
- 11 Taft M. Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes*, 1994, 9(3): 271—294
- 12 北京语言学院语言教学研究所. 现代汉语频率词典. 北京: 北京语言学院出版社, 1986
- 13 上海交通大学汉字编码组, 上海汉语拼音文字研究组. 汉字信息字典. 北京: 科学出版社, 1988

- 14 Burani C, Salmaso D, Caramazza A. Morphological structure and lexical access. *Visible Language*, 1984, 18: 342—352
- 15 Laudanna A, Burani C, Cermele A. Prefixes as processing units. *Language and Cognitive Processes*, 1994, 9(3): 295—316
- 16 Burani C, Laudanna A. Units of representation of derived words in the lexicon. In: Frost R, Katz L (ed), *Orthography, phonology, morphology, and meaning*. Amsterdam: North-Holland, 1992
- 17 Zhou X, Marslen-Wilson W. Words, morphemes and syllables in the Chinese mental lexicon. *Language and Cognitive Process*, 1994, 9(3): 393—422
- 18 Zhou X, Marslen-Wilson W. Morphological structure in the Chinese mental lexicon. *Language and Cognitive Processes*, 1995, 10(6): 545—600
- 19 Taft M, Huang J, Zhu X. The influence of character frequency on word recognition response in Chinese. In: *Advances in the study of Chinese language processing (Vol. 1)*. Department of Psychology, National Taiwan University, 1994

THE ROLES OF SURFACE FREQUENCIES, CUMULATIVE MORPHEME FREQUENCIES, AND SEMANTIC TRANSPARENCIES IN THE PROCESSING OF COMPOUND WORDS

Wang Chunmao

(Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101)

Peng Danling

(Department of Psychology, Beijing Normal University, Beijing, 100875)

Abstract

The present experiments focused on the structure of access representation. Surface frequencies, cumulative morpheme frequencies, and semantic transparencies of Chinese two-character words were controlled. When transparencies were balanced, the cumulative morpheme frequency effect was found only in high surface frequency words but not in low frequency words. In the next experiment surface frequencies were controlled and cumulative morpheme frequencies and transparencies were varied. The responses of words with high cumulative frequencies were faster than that with low cumulative frequencies in transparent words. In opaque words, however, this response tendency was reversed, namely, the responses of words with high cumulative frequencies were slower than that with low cumulative frequencies. The two experiments implied that both word and morpheme units received activation from visual input and there were connections between them. These connections were positive between transparent words and their morphemes, negative between opaque words and their morphemes, and would be enhanced with surface frequencies.

Key words access representation, surface frequency, cumulative morpheme frequency, semantic transparency.