

关于语料库搜索表达“责备”的形式手段过程的一些看法

——以“сам + небось + V”构式为例

Р.И. 施姆拉克¹ 郭鸣鹤²

(1 浙江大学外国语言文化与国际交流学院, 杭州 310058; 2 长春理工大学外国语学院, 长春 130013)

提 要: 本文主要探讨语料库搜索表达责备的形式手段的过程。在现代著作中, 责备被认为是一种必须结合语境理解的现象。然而, 这种说法忽略了俄语语篇中对于母语是俄语的人而言存在无需语用信息理解的责备, 这种类型的责备表明了存在抽象的责备构式, 通过表达责备的形式手段在特定的语篇层面上实现。与系统语言手段相比, 表达责备的形式手段很难通过直接观察和随后的语言分析来识别, 因为它们的使用比例要低很多。语料库语言学方法使我们能够解决这个问题。俄语国家语料库(НКРЯ)的搜索引擎可以输入复杂的搜索条件, 不仅包括词素, 还包括语法、语义和句法特征, 这让我们可以得到具有异质构造的责备构式。我们认为90%以上搜索结果带有责备语义可证明该形式手段的规律性, 可以认为它是表达责备的形式手段。本文探讨语料库搜索过程各个阶段所涉及的一些问题, 并提供了一种语料库搜索的试验算法, 用于搜索表达责备的形式手段。

关键词: 责备; 俄语国家语料库; 语料库语言学; 语料库搜索过程; 构式语法

中图分类号: H35 **文献标识码:** A

1 引言

最近几十年的一些著作将“责备”作为研究对象(Давыдова 2003; Каразия 2004; Федорова 2004), 虽然这些著作采用不同的方法解决不同的任务, 但共同点都认为语境是责备的前提, 责备是完全依赖语境的。但我们的研究发现, 这种看法是站不住脚的, 因为存在大量无需语境即可理解责备的例子。例如: *Прямо бомж какой-то. А сам небось коттеджик строит метров на восемьсот.*¹ (就像个流浪汉。但怕是正在建八百平方米的别墅。) 其中 *а сам небось коттеджик строит* 是完全可以表达责备的, 并且没有责备之外的语义, 即使去掉用作对照的 *прямо бомж какой-то* 以及修饰宾语 *коттеджик* 的定语 *метров на восемьсот*, 责备语义也不会改变。与此同时, 某人可能正在建造一栋小别墅的事实在语义上是中立的, 并不能表达责备。因此, 我们能够做出两个合理的假设, 作为我们研究的基础:

- 1) 语境不是理解责备语义的强制性条件;
- 2) 俄语语篇中应该有一些无需结合语境即可表达责备的形式手段。

这些论断并不与认为语境是形成责备的最重要手段的观点相抵触, 只是证实了无需语用信息的形式手段表达责备的可能性。例如: *Тожже мне — старший бронемастер драной армии... Сам небось драпал, все бросил, аж сюда додрапал, от своих же небось солдат спасался...* (什么破烂军队的高级装甲军师……他大概自己逃跑了, 抛下了一切, 扔下了自己的战士, 跑了那么远……) 一方面, 这里的责备显然是通过广泛的语境手段来实现的, 即

使没有读过这本书也能知道事件发生在军队溃败和士气低落的背景下，而 старший бронемастер 作为一个高级军官却败坏自己的名声，从军营中逃跑。语境加强了说话人明显的不赞同态度，通过加强语气词 же 和 аж、старший (бронемастер) — драной (армии) 的对照、未完成语调、省略号以及大量表示消极含义的动词 драпал、бросил 和 спасался 体现，语境直接涉及了责备的形成。另一方面，应该指出的是，сам небось драпал 本身即可表达责备。表面上该句通过具有贬义的动词 драпать 表达责备，即使用相同词义的任何其他中性动词来替换，责备语义也将保持不变，例如 сам небось бежал。实际上在 а сам небось коттеджик строит 和 сам небось драпал 两个看似无关的责备构式中有亲属关系的存在。构式语法 (Fillmore 1988, 1992; Goldberg 1995, 2003; Рахилина 2010) 中这两个例子并不是单独的现象，而都是责备构式 сам + небось + V 在语篇中的实现形式。俄语国家语料库所有分库的词汇语法搜索显示，该构式的 161 个搜索结果中，150 个带有责备语义²，比例为 93.16%。表 1 显示了语料库中 сам + небось + V 构式搜索结果的部分示例。

表 1. 语料库中 сам + небось + V 构式搜索结果部分示例

编号	示例
1	— Да это вы, мужчины, так рассуждаете, — говорила дама, не сдаваясь и оглядываясь на нас, — сами себе дали свободу, а женщину хотите в терему держать. Сами небось себе все позволяете. (“你们男人就是这么想的，”女士没有认输，环顾四周说，“给自己自由，却想把女人关在楼里。你们怕是什么都允许自己做。”)
2	— Так Бог сделал, чтобы мы умирали. — А сам, небось, никогда не умирает! Декабрь 18. Опять у Анненкова, пишу коротышку об Арцыбашеве. (上帝就是这样，让我们去死。他自己可能永远都不会死！12月18日。又在安年科夫这里，写关于阿尔茨巴舍夫的短文。)
3	Сами, небось, бифштексы жрут, а нам — треска. (他们大概自己大吃牛排，却给我们吃鲑鱼。)
4	Сам небось паклю себе взбил, девушкам на погибель... (他大概自己做密密麻麻的蓬松头发，却让女孩们受罪……)
5	Протираешь овощи — протирай, но не до дыр! Сам небось харчо жрет! (蔬菜擦丝就擦丝，但不要擦破！他自己怕是狼吞虎咽地喝牛肉汤！)
6	— А сами небось сегодня же побегите со своими стишками в «Одесский листок». Знаю я вас, молодых поэтов! Не обманете! (蔬菜擦丝就擦丝，但不要擦破！他自己怕是狼吞虎咽地喝牛肉汤！)
7	Все про твою полчку интересовался. А сам небось миллионы гребет (他一直对你地薪水感兴趣，而自己想必能挣上百万。)
8	— Ты же у нас тоже всех разумней, как послушать. А сам небось первым штаны замарал... (要是听你的，你比我们所有人都聪明。但你自己怕是第一个裤子弄脏地……)
9	— Сам-то небось кончал какой-нибудь оренбургский неплюевский... (自己大概是什么奥伦堡涅普留耶夫毕业的吧……)
10	А сама, небось, наверху себе пристроила еще одну душевую. (她想必是又在楼上给自己建了个淋浴房。)

上文表明存在所谓“固有的”责备，这些责备在语篇中无需语境就可以发挥作用³，是抽象责备构式的语篇实现手段，通过表达责备的形式手段在具体层面上表现出来。

形式语言手段的一个不可或缺的特性是规律性。一般情况下，通过直接观察和随后的语言分析来确定系统形式手段的规律性并不难。但是，基于语篇中表达责备的形式手段的抽象责备构式形象化需要一种特殊的方法，因为责备现象并不属于系统语言手段，而且与系统语言手段相比，责备在语篇中出现频率较低。因此需要一个结合构式语法原则、形式手段规律

性原则和语料库工具的方法，原因如下：

1) “所见即所得”(Goldberg 2003: 219)，构式语法使我们能够解释语篇中固有的责备，它们是抽象构式的实现手段；

2) 如果构式在俄语国家语料库词汇语法搜索中 90% 以上搜索结果带有责备语义，则可以证实该构式具有规律性，是表达责备的形式手段；

3) 使用语料库进行构式研究的想法已经被证明是很有有效的。从突破性著作《Collostructions: Investigating the interaction of words and constructions》(Stefanowitsch, Gries 2003)发表以来，语言学家越来越多地使用语料库工具进行构式研究(Ли 2014; Петрухина, Юйе 2019)。我们的研究也可以使用语料库，因为责备构式是包括词素、语法、语义、句法和其他特征的具有异质构造的构式。使用直接观察法在语篇中识别这种构式是非常困难甚至根本不可能的，而语料库搜索引擎为研究者提供了工具，用于创建与所需构式完全相同的复合搜索条件。

我们开发了一个试验程序，用于搜索俄语语篇中表达责备的形式手段。本文将介绍这一试验程序的各个步骤，以及测试过程中出现的问题和任务。

2 初步收集材料阶段

进行语料库搜索前需要准备待搜索构式清单，这一过程在很大程度上受所谓的“第一步”问题(搜索前由智力和心理活动引起的因素)影响。但是，初步收集材料阶段的主观性在一定程度上可被系统性现象所抵消。最近几十年来的研究者已经注意到语篇中存在的系统性现象。共现词素(collexemes)，即被特定构式吸引的词素的提出是研究者注意到语篇中系统性现象的第一个信号(Stefanowitsch, Gries 2003: 215)。此外，术语“话语‘表面’标记(маркеры на «поверхности» речи)”(Иссерс 2008: 133)、“语篇标记(дискурсивные маркеры)”(Шилихина 2015: 120)以及“关键短语或文本标记(ключевые фразы или текстовые маркеры)”(Дементьев, Степанова 2016: 57)也与所描述的现象有关。在我们的研究中，关于语料库搜索潜在的责备构式，我们认为可以讨论**第一近似标记**和**第二近似标记**。

责备标记是初步收集材料阶段的主观单位，是语篇的信号段(词素、词素组合和/或文本的其他特征)，帮助研究者怀疑语篇中存在潜在的责备构式。第一近似标记和第二近似标记是指标记的责备特征性，即标记可以指出研究者主观感知责备的频率。

直观看来，语气词 *небось* 可能是责备构式的共现词素，带有该词的示例经常表达责备，例如：*— Да уж в этом-то вы, должно быть, разбираетесь, — сказала я, нехорошо улыбаясь. — Небось уже по уши налили этим ликёром?* (“这个你们应该是知道地，”我坏笑着说，“你们恐怕甜酒喝高了吧？”)句中语气词 *небось* 的责备特征性实际上并不高。2020年5月31日俄语国家语料库主库中 *небось* 的精确搜索结果显示，在 2239 个文件的 6112 个示例中，至少半数不属于责备语义场，例如：*— Что ему твоя курятина! — снова загредел дядя Сандро и с хрустом перекусил сочное оперенье зелёного лука, — он небось хочет послушать что-нибудь из того, что случилось со мной.* (“他才不关心你的鸡肉！”桑德罗叔叔再次大喊起来，咬了一口脆生生又多汁地葱花，“他怕是想听听发生在我身上的一些事情。”)这种情况下，如果研究人员凭直觉假设词素与其形成责备的能力之间存在联系，则有理由将词素识别为第一近似标记，并进行进一步的搜索。

在俄语国家语料库词汇语法搜索中输入语气词 *небось* 和代词 *сам* 的组合(*сам* 与 *небось* 的距离为 1 至 3)，共得到 222 个示例，其中只有 26 个不带有责备语义(基于 2020 年 5 月 31 日搜索结果)，例如：*Сам небось не дурак, понимаешь, в какой переплет попал.* (你看了不傻，知道处境多么糟糕。)因此，我们可以认为 *сам + небось* 的组合是具有高度责备特征性(88.28%)的责备第二近似标记⁴。

通过对第二近似标记 сам + небось (сам 与 небось 的距离为 1 至 3) 搜索结果的分析, 我们可以总结初步收集材料阶段, 做出 сам + небось + V (сам 与 небось 的距离为 1 至 3, небось 与 V 的距离为 1 至 3) 是潜在责备构式的假设。证明这种构式是语篇中表达责备的形式手段只能在语料库搜索阶段进行。

3 语料库搜索阶段

图 1 展示了基于 сам + небось + V 构式 (сам 与 небось 的距离为 1 至 3, небось 与 V 的距离为 1 至 3) 通过语料库搜索表达责备的形式手段的试验算法。

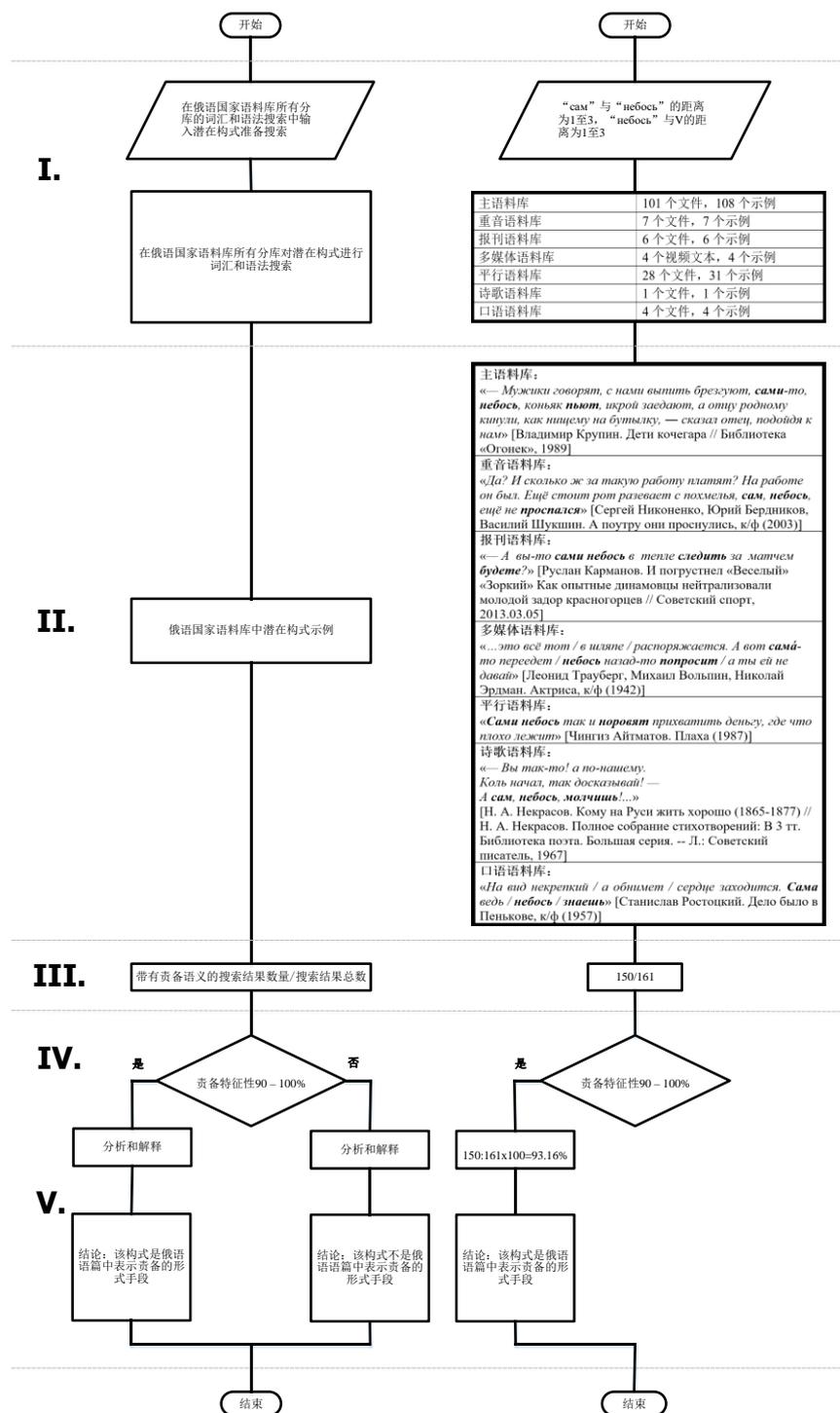


图 1. 基于 сам + небось + V 构式 (сам 与 небось 的距离为 1 至 3, небось 与 V 的距离为 1 至 3) 通

过语料库搜索表达责备的形式手段的试验算法

该算法可以统一在初步收集材料阶段发现的所有潜在责备构式的搜索程序,从而大大提高研究结果的客观性。作为“第一步”问题的初步收集材料阶段的局限性对后续语料库搜索阶段影响并不大,“最后一步”问题才是最关键的,这与在理解获得的客观语料库数据时不可避免的主观性有关(详情参见 Шмурак 2021)。由于语料库搜索的特点,证明潜在构式规律性的统计数据不能精确计算,我们只能进行相对的统计,分析统计趋势,而不是像自然科学和技术研究通常使用的“精确数字”统计。以下是我们在研究过程中这一问题的一些表现。

1) 判断示例是否带有责备语义对于语料库责备构式分析非常重要,也是容易出错的。首先,研究者在判断每个责备的语篇实现手段时是主观的。一些示例的责备语义不明显,例如: *Сам, небось, знаешь, как твоего брата проучивают!* (你看来知道你兄弟是怎么被教训的!)还有一些示例表现其他类型的言语侵犯,例如: *Сам-то, небось, унтеров не любил? Как вы нас промеж собой называли — шурами?* (你看来不喜欢士官吧?你们叫我们什么?痞子?)句子表现为威胁语义。上述情况都增加了统计的误差。无论选择材料的理论标准,对母语人士的大规模调查,语言学家团队验证还是其他此类措施都无法有效解决判断示例是否带有责备语义时的主观性问题,因为其根源远远超出了语言学的范畴,属于无意识过程,研究者在研究语料库搜索结果过程中的智力投入及其固有观点也包括在这一过程中。这个问题可以用 Ronald Coase 的经典名言来描述——“如果折磨数据足够长的时间,它就会承认”(转自 Klein 2001: 84)。

2) 搜索参数对统计信息的影响是巨大的,搜索参数的最轻微变化将导致搜索结果的显著变化。例如,减少 сам + небось + V 构式中各部分间距离(сам 与 небось 的距离为 1 至 3, небось 与 V 的距离为 1 至 3 改为 сам 与 небось 的距离为 1 至 1, небось 与 V 的距离为 1 至 1)后,俄语国家语料库所有分库中搜索结果数量从 161 个减少到 61 个。如果将搜索参数改为倒装构式(该倒装构式可视为原构式的变体),即 небось 与 сам 的距离为 1 至 3, сам 与 V 的距离为 1 至 3,则搜索结果数量将减少到 124 个。逻辑上的解决方案是统一搜索查询规定,但因各种构式的异质性和唯一性,这是不可能的。

3) 在语篇实现手段层面上,责备构式的特点是存在不同变体,将这样的责备构式视为单独构式还是变体仍然存在争议,研究者做出的决定自然会改变统计数据。两种构式(或两种变体)сам + небось + V 和 сам + небось + Adj 同样都表达固有的责备。仔细研究示例—*А сами небось женатые, — сказала худая гостья.* (“他们想必结婚了。”瘦小的客人说。)和 *Сам небось женился, а других отговариваешь.* (你想必结婚了,却劝阻别人。)很明显,它们只在谓语的类型上有所不同,并且可以总结为一个构式——сам + небось+谓语。

4) 统计数据还受到手动处理大量语料库输出搜索结果(几千到几万个示例)技术复杂性的限制,有大量搜索结果的固有责备构式虽然少见,但并不是没有。例如,“не + стыдно + 第二人称代词”有 2666 个搜索结果(基于俄语国家语料库 2020 年 1 月 12 日数据)。

5) 语料库中同一构式在搜索参数相同的条件下搜索结果数量可能会随时间而改变。对语料库一些责备构式六个月的监测结果表明,随着时间变化,搜索结果的数量并不是不变的。语料库文本逐渐增加,语料库规模扩大,因此能观察到搜索结果增加的趋势。此外,软件解决方案、技术错误以及错误的搜索算法也可以引起搜索结果数量的变动。

上述问题是否是俄语语篇中固有责备运作方式系统化过程中不可逾越的障碍?相对统计数据是否可以作为通过语料库识别表达责备的形式手段的有效工具?我们对第一个问题的答案是“当然不是”,对第二个问题的答案是“当然是”。我们研究的目的是将俄语语篇中固有的责备形象化,而不是用数学方法精确地计算语料库文本中此类责备的数量,因此可以认为在理性考虑上述误差的情况下,我们的统计数据是客观的。

4 结束语

完成本文过程中, 我们为自己设定了一个目标, 即证明俄语语篇中存在固有的责备, 并描述作为验证该观点经验方法的语料库试验搜索程序。研究结果如下:

1) 尽管俄语语言学传统认为责备是一种必须结合语境理解的现象, 但实际上俄语语篇中有很多无需语境信息的固有责备。

2) 俄语语篇中固有责备的存在意味着存在抽象构式, 这些抽象构式可以通过在特定语篇层面上表达它们的形式手段来形象化。

3) 由于形式手段的主要特性是规律性, 根据我们的研究, 在形象化语篇中表达责备的形式手段过程中, 如果一个构式 90% 以上的搜索结果带有责备语义, 则可证明该构式是语篇中固有的责备。

4) 由于表达责备的形式手段不属于系统形式手段, 因此我们开发了一种试验方法, 使我们能够凭借经验将俄语语篇中固有的责备形象化。该方法结合了“构式语法”以及形式手段具有规律性的原则, 同时利用俄语国家语料库搜索引擎, 创建用于搜索具有复杂异质构造的潜在责备构式的复合搜索条件。

5) 基于 сам + небось + V 的例子按步骤展示了将表达责备的形式手段形象化的过程, 包括初步收集材料阶段和语料库搜索阶段, 以及过程的局限性。

我们认为, 本文的数据可作为俄语语篇中存在固有责备的证据, 并展示了语料库在未来进行此类研究的潜力。

附注

1 本文所引用例句均来自俄语国家语料库, 下同。

2 在 11 个没有责备语义的示例 (比例为 6.84%) 中, 语气词 небось 均表示其“原始的”认知情态, 例如: Оставалось беспокоиться и ждать — сам небось мне скажет. (只能担心和等待了, 他可能会告诉我。)

3 此处和下文中术语“固有的责备”均表示该含义。

4 应当指出, “代词+语气词”的组合经常作为责备第二近似标记。

参考文献

[1] Давыдова Т. А. Речевой акт упрека в английском языке: Дисс. ... канд. филол. наук[D]. Иркутск: Иркутский государственный лингвистический университет, 2003.

[2] Дементьев В. В., Степанова Н. Б. Корпусные методы в исследовании речевых жанров: проблема ключевых фраз[J]. Вестник РУДН. Серия: Лингвистика, 2016(3).

[3] Иссерс О. С. Коммуникативные стратегии и тактики русской речи[M]. Москва: Издательство ЛКИ, 2008.

[4] Каразия Н. А. Прагмалингвистическое исследование акта упрека в контексте современной американской речевой культуры: Дисс. ... канд. филол. наук[D]. Петропавловск-Камчатский: Камчатский государственный педагогический университет, 2004.

[5] Ли Чжухонг. Корпусный анализ форм будущего времени в русском языке: статистика и коллокационные закономерности[J]. Вестник Московского университета. Серия 9. Филология, 2014(6).

[6] Петрухина Е. В., Юйе Шэнь. Состав, семантика и частотность возвратных конструкций с дативным субъектом в русском языке (на материале Национального корпуса)[J]. Вестник Московского университета. Серия 9. Филология, 2019(2).

[7] Рахилина Е. В. Лингвистика конструкций[M]. Москва: Азбуковник, 2010.

[8] Федорова А. Л. Речевая стратегия упрека: лингвокогнитивный подход (на материале немецкого,

- английского и русского языков): Дисс. ... канд. филол. наук[D]. Уфа: Башкирский государственный университет, 2004.
- [9]Шилихина К. М. Изучение дискурсивных маркеров методами корпусной лингвистики[J]. Вестник ВГУ. Серия: Лингвистика и межкультурная коммуникация, 2015(3).
- [10]Шмурак Р. И. К уточнению понятия упрека с помощью корпусных инструментов[J]. Известия Российской академии наук. Серия литературы и языка, 2020(3).
- [11]Шмурак Р. И. Формальные средства выражения упрека в русском языке: к процедуре корпусного поиска (на материале конструкции «сам + небось + V»)[J]. Вестник Московского университета. Серия 9. Филология, 2021(2).
- [12]Fillmore C., Kay P. Construction grammar course book[M]. Berkeley: University of California, 1992.
- [13]Fillmore C., Kay P., O'Connor K.T. Regularity and idiomaticity in grammatical constructions: the case of let alone [J]. Language, 1988(64).
- [14]Goldberg A. E. Constructions: A Construction Grammar Approach to Argument Structure[M]. Chicago: University of Chicago Press, 1995.
- [15]Goldberg A. E. Constructions: A new theoretical approach to language[J]. Trends in Cognitive Sciences, 2003(5).
- [16]Klein D. B. A Plea to Economists Who Favour Liberty: Assist the Everyman[M]. London: Institute of Economic Affairs, 2001.
- [17]Stefanowitsch A., Gries S. T. Collocations: Investigating the interaction of words and constructions[J]. International Journal of Corpus Linguistics, 2003(2).

Remarks on the Procedure of Corpus Search for Formal Means of “Reproach”

— Based on the Construction “sam + nebos' + V”

R. I. Shmurak¹, Guo Ming-he²

(1 School of International Studies, Zhejiang University, Hangzhou 310058; 2 School of Foreign Languages, Changchun University of Science and Technology, Changchun 130013 China)

Abstract: This article is concerned with the procedure of corpus search for formal means of reproach. In modern works, reproach is considered as a phenomenon, which needs the context as a sine qua non condition. This understanding, however, is not consistent with the possible situation for discourse utterances that are perceived by Russian native speakers as reproaches, regardless of the pragmatic information "dissolved" in the context. The existence of such reproaches in the Russian language suggests the existence of abstract reproach constructions implemented through formal means of reproach at the level of discourse. In contrast to formal language means, formal means of reproach are difficult to identify through direct observation and subsequent linguistic analysis, since their frequency of use is disproportionately less. The use of corpus linguistics methods could solve this issue. The search engine of the Russian National Corpus allows complex query creation containing not only lexemes, but also grammatical, semantic and syntactic characteristics. This provides the possibility to get reproach constructions with their composite heterogeneous structure. A potential construction, which has more than 90% of the search results falling under the semantics of reproach, demonstrates the regularity

inherent in formal language means, and, in our opinion, should be evaluated as a formal means of reproach. The article explores a number of issues related to all stages of the corpus search procedure, as well as proposes an experimental corpus search algorithm for formal means of reproach.

Key words: reproach; Russian National Corpus; corpus linguistics; corpus search procedure; construction grammar

作者简介: P.H. 施姆拉克, 浙江大学外国语言文化与国际交流学院副教授, 研究方向: 语料库语言学, 句式语法, 比较语言学。郭鸣鹤 (1979-), 女, 吉林梅河口人, 硕士, 长春理工大学外国语学院讲师, 研究方向: 外国语言学及应用语言学。

收稿日期: 2022-03-23

[责任编辑: 李 侠]