

基于 NLT 检索系统的词语搭配对比

——以指向动词“めざす”和“ねらう”为例

和 佳

(云南大学旅游文化学院 外语系, 云南 丽江 674100)

摘要:区分日语指向动词“めざす”和“ねらう”可以利用 NLT 检索系统分析词语搭配。从共现频数、MI 值和 LD 值看,两者的显著搭配名词有“向上”“扩大”等。与“めざす”显著搭配的名词有“实现”“取得”等,体现了和动作主体的关联性和志向性,且几乎没有贬义词。与“ねらう”显著搭配的名词有“效果”“命”等,体现了限定性和加害性,包括中性、褒义和贬义词。利用 NLT 检索系统了解近义词的词语搭配异同能够为学习者提供一些参考。

关键词:めざす;ねらう;NLT;共现频数;MI 值;LD 值

中图分类号:H36 文献标识码:A 文章编号:1674-5639(2017)01-0127-06

DOI:10.14091/j.cnki.kmxyxb.2017.01.020

A Study on Collocation Based on the NLT Japanese Online Retrieval System

——Taking Japanese Direction Verbs “めざす” and “ねらう” as Examples

HE Jia

(Department of foreign languages, Culture and Tourism College of YunNan University, Lijiang, Yunnan, China 674100)

Abstract: In order to distinguish Japanese direction verbs “めざす” and “ねらう”, the Online Japanese Retrieval System of NLT can be used to the collocation. From frequencies of collations, MI values and LD values from each word, it can be found that the nouns collocate with words of “向上”“扩大”. The nouns collocating with “めざす” are “实现”“取得”, which reflecte the relevance and ambition from the subjects, and there are hardly any derogatory meanings. The nouns collocating with “ねらう” are “效果” and “命”, which reflecte limitation and tortuous action and include neutral words, commendatory terms and derogatory words. NLT retrieval system is used to understand the similarities and differences about collocation of synonym and can provide a certain reference for learners.

Key words: めざす;ねらう;NLT;frequency;MI score;LD score

传统语言分析的方法是以定性分析为基础的,主要依靠人们对语言现象的主观分析与评价。而基于语料库的方法则是定性定量相结合的方法,可以利用语料库得出的定量信息对语言现象进行分析、评价。语料库语言学是以语料库为基础对语言进行客观描述,研究语言运用的新手段。严格来说,语料库语言学不是一个新的学科,而是一种新的研究方法。^[1]换言之,语料库语言学研究方法已成为现代语言学研究的一个重要研究手段。

日语中存在大量的近义词,日语的近义词对学习者来说习得困难。特别是在交流中起重要作用的

动词,因为其包含很多意思,所以要完全正确地使用动词的近义词不是一件容易的事。比如,日语中属于近义词,且使用频率极高的“思う”和“考える”,经常会出现误用。词汇是许多第二语言学习者的主要障碍之一。对第二语言学习者的研究^[2]发现,词汇错误比语法错误多。教师能否让学习者准确掌握近义词意义的异同,是词汇指导中关系到日语能力提升的重要事项。^[3]

词语搭配被认为是一个非常重要的概念。最早提出了词语搭配“collocation”一词的是 Firth (1957:

收稿日期:2016-10-15

作者简介:和佳(1981—),女(纳西族),云南丽江人,讲师,硕士,主要从事日语语言研究。

195)。^[4]词语搭配一般表示词组关系,即词和词的组合关系,是根据一般语法规则共起的词和词的惯用关系。^[5]词语搭配是最为活跃的研究领域,处于现代语料库语言学研究的中心地位。^[6]

很多研究都利用语料库研究词语搭配,以便更好地区分近义词。比如,部田(2009年)^[7]、川口(2012年)^[8]、木山(2013年)^[9]、佐々木(2014年)^[10]等,都是利用语料库对句型、助词、动词词语搭配等进行了研究。但是有关“めざす”和“ねらう”研究的资料不多,故本文以它们为研究对象,利用日语语料库的检索系统 NLT 考察日语指向动词“めざす”和“ねらう”的词语搭配,试图探讨两者的异同。

一、NLT 检索系统及相关数值

NLT 全称为“NINJAL - LWP for TWC”,是以“筑波ウェブコーパス”(Tsukuba Web Corpus,简称 TWC)为母体的检索系统。该语料库系筑波大学从日语网站收集语料构建而成,容量约为 11 亿词。该检索系统于 2013 年公开,检索利用国立国语研究所和 Lago 言语研究所共同开发的语料库检索系统 NINJAL - LWP (NINJAL - Lago Word Profiler)。此外,以日本国立国语研究所研制开发的容量约为 1 亿词的大型语料库“現代日本語書き言葉均衡コーパス”(Balanced Corpus of Contemporary Written Japanese,简称 BCCWJ)为母体的检索系统“NINJAL - LWP for BCCWJ”,简称 NLB (<http://nlb.ninjal.ac.jp>)也属于利用相同系统的检索软件。

利用 NLT 检索系统可以用来全面检索名词、动词、形容词、连体词、副词以及语法标记的搭配关系。(网址为:<http://corpus.tsukuba.ac.jp>)。所使用的形态素解析器为:MeCab 0.98 + IPA 词典 2.7.0,系受关系解析器为 CaboCha 0.60。

从语料库的规模来看,“現代日本語書き言葉均衡コーパス”采用严密的统计方法采集数据,所以会受到时间、资金方面的制约,相比之下,“筑波ウェブコーパス”利用网络收集文本,则没有这方面的制约。和容量 1 亿词的语料库相比,容量 10 亿词的语料库包含了更多有用的语言信息。^[11]

参考井本亮(2013年)^[12]、赤瀬川(2014年)^[13]的研究方法,本文利用共现频数、MI 值和 LD 值,分析词语搭配情况。频数指要统计的语料库中所有类符出现的频率,对频数的统计分析是语料库研究的一个重要方面。频数 10 以上得不到结果的可以设定为频数 20 以上。^{[14][19]}

共现词显著性的测量方法通常有两种:MI 值和 T 值^①。这两种方法都是通过比较共现词的观察频数和期望频数的差异来确定搭配序列在语料库中出现概率的显著程度。^[15]MI 值(Mutual Information Score,互信息值)表示的是互相共现的两个词中,一个词对另一个词的影响程度或者是一个词在语料库中出现的频数所能提供的关于另一个词出现的概率信息。^[16]MI 值的计算公式为:

$$MI(x, y) = \log_2 [P(x, y) / P(x) \times P(y)] = \log_2 [F(x, y) \times N / (F(x) \times F(y))].$$

基于语料库的词语搭配研究中通常把 MI 值等于或大于 3 的词作为显著搭配词。^[17]

NLT 检索系统中的另外一个重要数据是 LD 值。LD 值是 logDice 系数,取的是 Dice 值(语言表现中分析共起关系常用的统计值之一)的对数。用于词语搭配的统计处理。从共起频数、节点词的频数、搭配词频数等 3 个变数得到的统计值,按照降序排列,可以得到比 MI 值更接近单纯频数的结果。^[14]20LD 值越大,表明词语搭配约显著。^{[14][33]}

NLT 检索系统的 logDice 以 SketchEngine 在线语料库索引工具为基准。LD 值的计算公式为^[18]:

$$\logDice = 14 + \log_2 (2 \times |X \cap Y| / (|X| + |Y|)).$$

二、两者的异同对比

(一) 基于词典的对比

《广辞苑》^[19]对两者的解释为:

めざす:1. めあてとしてその方を見る。
太平記三「一・すとも知らざる暗き夜に」

2. 目標とする。めがける。ねらう。「北極点を一・す」「政治家を一・す」

ねらう:1. あるものに目をつけ、手に入れ

^①T 值是根据假设检验中的 t 检验计算得来的。假设检验主要通过检验某一样本的平均数与正太分布总体的平均数之间的差异是否显著来判定该样本取自总体的可能性有多大,或者说二者之间的差异是否由偶然性造成。

ようとして機会を待つ。万八「この岡にを鹿踏み起て窺一・ひも右もすらく君ゆゑにこそ」。「すきを一・う」

2. 目標に命中させようとした構える。殺そうとして機会をうかがう。平家一「九郎大夫の判官を射落さんと一・はれけれども」。「的を一・う」あることがらを目標とする。「俗受けを一・う」

通过上面的定义,笔者将两者的共同意义归纳为“以…作为目标”。区别是:“めざす”还可以指朝目标看,“ねらう”主要强调为了实现目标而作准备。可以指看到某物,等着到手的机会或者为了命中目标准备好;寻找杀人的机会。但是我们发现,字典中对“めざす”的解释使用了“ねらう”。《日本国語大辞典》等词典的解释,对“ねらう”的解释很详细,但是查找“めざす”出现的解释是“めがける、ねらう”。查找“ねらう”,出现的解释是“めざす”。^[20] 词典的描述就是“定义的来回兜圈子”^[21]。除了国语辞典以外,一些日英词典,比如研究社《新和英大辞典》、小学館《プログレッシブ英和中辞典》的 aim,白水社《現代スペイン語辞典(改訂版)》的 pretender 的译语,把“めざす”、“ねらう”写

在一起。^[22]

这样的解释不利于学习者掌握它们的意义及区别,以下通过 NLT 检索系统再来看两者词语搭配的同异。

(二)基于 NLT 检索系统的对比

在 NLT 检索系统输入“めざす”,系统显示的频数为 157 443。具体类型包括:名词+助词+めざす;名词+复合助词+めざす;めざす+名词;めざす+助动词;めざす+复合动词;近接动词^①+めざす;めざす+形容词;副词+めざす;形容词连用形+めざす;未分类等 10 类。“ねらう”显示的频数为 38 059,具体类型与“めざす”完全相同。本文只考察两者“名词+助词+节点词”这一类型的用法。其中,助词包括“が、は、も、の、を、に、へ、で、と、から、まで、より”等,本文选取频数最高的“を”为研究对象。结果显示,“めざす”和“ねらう”的频数分别为 157 443 和 25 759。

分别点击“をめざす”、“をねらう”后,系统出现节点词的词汇搭配、共现频数、MI 值和 LD 值。笔者将共起频数排在前 30 的 MI 值和 LD 值分别进行统计,总结为下表 1 和表 2。

表 1 节点词“めざす”共现频数前 30 位的 MI 值和 LD 值

序号	をめざす	频度	MI	LD	序号	をめざす	频度	MI	LD
1	こと	11 471	3.41	5.28	16	企業	987	4.23	5.77
2	向上	3 471	8.47	9.05	17	家	978	3.82	5.44
3	実現	3 035	8.58	8.97	18	発展	952	6.79	7.25
4	化 ^②	2 619	5.68	7.21	19	者	873	1.69	3.51
5	づくり ^③	2 378	7.55	8.34	20	確立	811	8.66	7.50
6	一般 ^④	1 974	1.64	3.50	21	士 ^⑤	771	4.87	6.13
7	育成	1 749	8.18	8.30	22	師 ^⑥	758	6.17	6.81
8	地域 ^⑦	1 454	0.71	2.58	23	拡大	710	6.07	6.71
9	取得	1 450	7.64	7.94	24	何	706	6.84	6.98
10	作り	1 258	6.99	7.57	25	プロ	698	2.74	4.44

①近接动词是指该动词和前后 3·5 个词以内的共起动词的类型。

②指在名词、动词、形容动词等后面加“化”构成名词,如:三年後に実用化を目指すそうだ。

③指在名词后面接续“づくり”的形式,如:まちづくりを目指します。

④指除了人名、地名、组织名等以外的固有名词。

⑤指在名词后面接续“士”的形式,如:保育士を目指す皆さん。

⑥指在名词后面接续“師”的形式,如:看護師を目指す方へ。

⑦指一般的地名、国名等。

续表 1

序号	をめざす	频度	MI	LD	序号	をめざす	频度	MI	LD
11	社会	1 230	4. 19	5. 80	26	解決	686	6. 36	6. 79
12	合格	1 186	7. 75	7. 77	27	成長	655	5. 94	6. 59
13	構築	1 092	7. 93	7. 73	28	それ	599	5. 49	6. 31
14	アップ	1 052	7. 64	7. 62	29	完成	577	1. 65	3. 43
15	上	1 016	3. 41	5. 10	30	改善	519	6. 98	6. 69

注:节点词频数:157 443;搭配名词种类:6 341。

表 2 节点词“ねらう”共现频数前 30 位的 MI 值和 LD 值

序号	をねらう	频度	MI	LD	序号	をねらう	频度	MI
1	こと	947	2. 13	1. 70	16	利益	174	6. 22
2	効果	780	6. 67	6. 14	17	拡大	172	7. 13
3	一般 ^①	690	2. 45	2. 01	18	チャンス	169	7. 57
4	命	545	6. 96	6. 35	19	機会	159	5. 83
5	それ	358	3. 28	2. 83	20	優勝	158	9. 01
6	化 ^②	265	4. 70	4. 19	21	人	151	1. 07
7	ところ	245	3. 86	3. 38	22	魚	147	6. 43
8	人名	232	0. 62	0. 19	23	獲物	138	10. 65
9	とき	218	2. 68	2. 23	24	向上	134	6. 10
10	者	190	1. 82	1. 37	25	得点	130	8. 72
11	受け	186	10. 08	7. 51	26	の ^③	128	-0. 30
12	アップ	180	7. 41	6. 35	27	日	125	2. 10
13	そこ	177	4. 59	4. 05	28	何	124	2. 57
14	地域 ^④	176	-0. 01	-0. 44	29	タイミング	119	7. 07
15	隙	175	8. 32	6. 84	30	これ	115	1. 61

注:节点词频数:25 759;搭配名词种类:4 041。

从上面的表 1 和表 2 可以看出,“めざす”和“ねらう”有一些搭配相同的名词,如“こと”“向上”“化”“一般”“地域”“アップ”“者”“拡大”“何”“それ”等。只能和“めざす”共起的名词有:“取得”“作り”“社会”“構築”“家”“師”“プロ”“完成”“改善”等。只能和“ねらう”共起的名词有:“命”“者”“受け”“そこ”“隙”“チャンス”“機会”“魚”“獲物”“得点”“何”“タイミング”等。从搭配名词的种类看,“めざす”的搭配名词种类为 6 341,“ねらう”的搭配名词种类为 4 041,“めざす”的后续名词种类多于“ねらう”。从搭配相同的名词来看,“こと”与“めざす”和“ねらう”的共现频数都是最高的,但两者却相差甚远,分别为 11 471 和 947。

“向上”与“めざす”的共现频数 3 471,与“ねらう”的共现频数只有 134。“めざす”接续“一般”、“地域”的共现频数分别是 1 974 和 1 454,大于与“ねらう”的共现频数 690 和 176。此外,像“アップ”“者”“拡大”“何”“汗”与“めざす”和“ねらう”的共现频数分别为 1 052 和 180;873 和 190;710 和 172;706 和 124。可见,两者虽然搭配的名词相同,但是与节点词的共现频数是有差异的,总的来说,“めざす”搭配名词的共现频率明显大于“ねらう”的共现频率。

但是,仅根据共现频数的高低,研究者还无法确定每一个共现词是否为显著搭配词。还需要运用统计测量的方法,检验每一个共现词与节点词之间的

①指除了人名、地名、组织名等以外的固有名词。

②指在名词、动词、形容动词等后面加“化”构成名词,如:三年後に実用化を目指すそうだ。

③指在在动词、形容词、代词等后面接续“の”的形式,如:自動延長がないのを狙え!

④指一般的地名、国名等。

相互预见和相互吸引程度。^[16]所以,接着需要分析 *MI* 值和 *LD* 值的情况。

MI 值主要是通过测量共现词的非随机性来体现词语搭配的显著性。*MI* 值可以清楚表现共现词间的相互吸引程度,它可以帮助我们确定把哪些词作为节点词的可能搭配词而重点加以研究。^[16] *MI* 值作为统计指标之一,越具有代表性的词语搭配,其 *MI* 值越高。^[14] *LD* 值也是如此。鉴于 NLD 检索系统是以大型语料库为母体,故本文将 *MI* 值和 *LD* 值 ≥ 5 的词视为显著搭配。上表 1 中,“めざす”后接“こと”的频数最高,*LD* 值为 5.28,但 *MI* 值为 3.41,不视为显著搭配。“向上”“实现”“化”“づくり”的 *MI* 值和 *LD* 值均大于 5,且频数高,可视为显著搭配。“一般”的频数为 1974,但 *LD* 值为 3.5,但 *MI* 值只有 1.64,小于 3,不视为显著搭配。使用同样的方法检验表 2 后可以得知:与“めざす”和“ねらう”均属于显著搭配的名词有“向上”“アップ”“拡大”。与“めざす”显著搭配的名词有:“实现”“づくり”“化”“育成”“取得”“作り”“合格”“構築”“発展”“確立”“師”“何”“解決”“成長”“それ”“改善”等。与“ねらう”显著搭配的名词有:“效果”“命”“受け”“隙”“利益”“チャンス”“機会”“優勝”“魚”“獲物”“得点”“タイミング”等。

井上(1983年)^[20]提出“ねらう”和“めざす”的区别在于“意识到现在可能达成的对象”和“距离主体较远,或者意识到现在不可能达成的对象”。我们抽取 NLT 检索系统中的例句来看:

1. そういう効果を狙いたい。
想得到那样的效果。
2. 運動を長続きさせることを目指します。
目标是把运动坚持下去。

例句中“ねらう”修饰的名词“效果”不一定能够达到,而“めざす”修饰的名词“把运动坚持下去”,也不能断定是不可能实现的目标。因此,井上(1983年)的提法不够严谨。

从节点词“めざす”的显著搭配名词看,像“实现”“化”“づくり”“構築”等体现了和动作主体的关联性。如:

3. 夢の実現を目指します。
目标是实现梦想。
4. 5年後の法人化を目指す。

以 5 年后的法人化为目标。

5. 持続可能な社会の構築を目指す。

以构建可持续社会为目标。

像“師”“合格”“確立”表达了主体的志向。如:

6. 看護師を目指す理由は?

想成为护士的理由是?

7. 試験合格を目指します。

目标是考试及格。

8. まずお客様との信頼関係確立を目指します。

首先以确立与客户的信赖关系为目标。

从上面的例句可以看出,节点词“めざす”包含了经过时间和努力,尽力到达目标的意思。^[23]

从节点词“ねらう”的显著搭配名词看,像“魚”“獲物”“得点”等可归类为具体名词,体现了动词的限定性。如:

9. 小魚を狙ってダイビングします。

瞄准小鱼跳下去。

10. 獲物を狙っている様子。

瞄准猎物的样子。

11. TOEICで高得点を狙うならこの教材!

想在 TOEIC 取得高分,就用这本教材!

名词“命”的频度排第四位,*MI* 值和 *LD* 值均大于 5,接续“ねらう”指想要杀害,体现了加害性。如:

12. 兄から命を狙われるのも当然です。

哥哥想要他的命也是理所应当的。

从上面的例句可以看出,节点词“ねらう”的对象物既可以是像“夺命”这样的贬义词,也可以是像“優勝”“向上”这样的褒义词,或者“受け”“チャンス”这样的中性词。而“めざす”的显著搭配中几乎没有贬义词。

三、结语

利用日语大型语料库的 NLT 检索系统明确共现频数、*MI* 值和 *LD* 值,可以更好地掌握日语近义词的区分。本文以指向动词“めざす”和“ねらう”为研究对象,考察了接续名词的异同。通过对共现频数、*MI* 值和 *LD* 值的分析,可以得知与“めざす”和“ねらう”均属于显著搭配的名词有“向上”“アッ

プ”“拡大”。与“めざす”显著搭配的名词有：“实现”“づくり”“化”“育成”“取得”“作り”“合格”“構築”“発展”“確立”“師”“何”“解決”“成長”“それ”“改善”等，体现了和动作主体的关联性和志向性。这些名词几乎没有贬义词。与“ねらう”显著搭配的名词有：“効果”“命”“受け”“隙”“利益”“チャンス”“機会”“優勝”“魚”“獲物”“得点”“タイミング”等，体现了限定性和加害性。这些名词包括中性词、褒义词和贬义词。但是字典中对它们的解释显得较为笼统，学习者不容易掌握。利用NLT检索系统了解近义词的词语搭配异同，能够为日语学习者提供一定的参考。

[参考文献]

- [1] 顾曰国. 语料库与语言研究[J]. 现代语言学, 1998(1):1.
- [2] 江新. 词汇习得研究及其在教学上的意义[J]. 语言教学与研究, 1998(3):65-73.
- [3] 倉持保男. 日本語教育における類義語の指導[J]. 日本語学, 1986, 5(9):47-55.
- [4] RUPERT F J. Modes of meaning[C]//In Papers in Linguistics 1934-51. London:Oxford University Press, 1957:195.
- [5] 小池生夫. 応用言語学事典[M]. 東京:研究社, 2003.
- [6] 杨惠中. 语料库语言学导论[M]. 上海:上海外语教育出版社, 2002:82.
- [7] 部田和美. 授受動詞「ヤル・クレル・モラウ」文の意味分析—抽象的対象物を含む授受動詞文を中心に—[J]. 言語学論叢オンライン版, 2009(28):33-47.
- [8] 川口裕子. 日本語並立助詞「と」・「や」と英語冠詞に関する—考察—BCCWJデータに基づいて—[J]. 第一回コーパス日本語学ワークショップ予稿集, 2012:113-118.
- [9] 木山直毅. 日本語複合動詞「V直す」、「V返す」、「V戻す」の特徴[J]. 第三回コーパス日本語学ワークショップ予稿集, 2013:39-48.
- [10] 佐々木幸太. 主体変化動詞・状態動詞・天候動詞の「Vテイク形」・「Vテクル形」[J]. 日本文藝研究, 2014(66):1-21.
- [11] 今井新悟, 赤瀬川史朗, プラシヤント・パルデシ. 筑波ウェブコーパス検索ツールNLTの開発[J]. 第3回コーパス日本語学ワークショップ予稿集, 2013:199-206.
- [12] 井本亮. [現代日本語書き言葉均衡コーパス]にみられる副詞的修飾関係「赤くV」について[J]. 商学論集, 2013, 82(1):1-19.
- [13] 赤瀬川史朗, プラシヤント・パルデシ, 今井新悟. NINJAL-LWPの類義語比較機能[J]. 第6回コーパス日本語学ワークショップ予稿集, 2014:41-50.
- [14] 筑波大学, 国立国語研究所, Lago 言語研究所. NLT(NINJAL-LWP for TWC)ユーザマニュアルバージョン1.30[G]. 東京, 2015.
- [15] HUNSTON S. Corpora in Applied Linguistics [M]. Cambridge:Cambridge University Press, 2002.
- [16] 邓耀臣. 词语搭配研究中的统计方法[J]. 大连海事大学学报, 2003(40):74-77.
- [17] CHURCH K, Hanks P. Word association norms, mutual information and lexicography [J]. Computational Linguistics, 1990(16):22-29.
- [18] 今田水穂. 日本語特講Ⅱ[R]. 国立国語研究所コーパス開発センター, 2014:33.
- [19] 新村出. 広辞苑[M]. 第5版. 東京:岩波書店, 1998.
- [20] 井上次夫. 「[c3][c4]指向動詞の意味分析—めがける・ねらう・めざす—」[J]. 奈良教育大学国文, 1983(6):38-49.
- [21] 国広哲弥. 理想の国語辞典[M]. 東京:小学館, 1997:165.
- [22] 加藤理恵. 日本語教育の諸問題——類義語分析 指向動詞「めざす」と「ねらう」[J]. 第2回名古屋大学日本語教育研究会予稿集, 2004:30-33.
- [23] 金慶恵. 日本語動詞「ねらう」の意味論的考察—「めざす」との対照を通して—[J]. 日本近代学研究所, 2013(42):85-102.