



DOI: 10.22363/2312-9182-2018-22-3-560-580

A Corpus Investigation of English Cognition Verbs and their Effect on the Incipient Epistemization of Physical Activity Verbs

Sally Rice¹ and John Newman^{1,2}

¹University of Alberta
116 St. and 85 Ave., Edmonton, AB, Canada T6G 2R3

²Monash University, Australia

Abstract

In the spirit of NSM accounts that attempt to build up a language's full expressivity from a small set of lexical primitives, we have investigated the usage in English of basic verbs of ideation (*think, know*) and physical activity (*strike, hit, go, run*) as they take on new epistemic meanings and functions, all the while calcifying in their inflectional range. It is well known that certain verbs of cognition in English such as *remember, forget, and think* are grammaticalizing into pragmatic particles of epistemic stance and, consequently, 1st person singular (1sg) forms account for the majority of usages. Likewise, we have carried out systematic queries and hand-tagging of corpus returns and have found that many verbs and phrasal expressions, ideational or not, seem to be associated with rather narrow collocational patterning, argument structure, and inflectional marking in almost idiom-like and constructional fashion. Moreover, we find that expressions associated with 1sg and 2nd person "cognizers" are, to a large extent, in complementary distribution, giving rise to fairly strong semantic differences in how *I* and *you* "ideate". In this study, we demonstrate the extent of inflectional and collocational specificity for verbs of cognition and physical activity and discuss implications this lexico-syntactic idiosyncrasy has for cognitive linguistics.

Keywords: *cognition, corpus methods, collocation, lexico-syntax, inflectional categories, epistemization*

Корпусное исследование английских глаголов мышления и их влияния на начальную эпистемизацию глаголов физического действия

Сэлли Райс¹ и Джон Ньюман^{1,2}

¹Университет Альберты, Канада
116 St. and 85 Ave., Edmonton, AB, Canada T6G 2R3

²Университет Монаш, Австралия

Аннотация

С позиций подхода ЕСМ, который стремится представить всю экспрессивность языка с помощью небольшого набора лексических примитивов, мы исследовали употребление английских базовых глаголов мышления (*think, know*) и глаголов физического действия (*strike, hit, go, run*), которые принимают новые эпистемические значения и функции. Хорошо известно, что определенные глаголы мышления в английском языке, такие как *remember, forget, think*, могут использоваться

в роли прагматических частиц эпистемического содержания и, как следствие этого, форма 1 лица единственного числа является наиболее употребляемой. В результате проведения систематических исследований ручной аннотации корпуса мы обнаружили, что для многих глаголов, как и фразовых выражений, характерна достаточно узкая сочетаемость, аргументативная структура и некоторые черты идиоматичности. Более того, нами обнаружено, что в выражениях, где глаголы мышления употребляются с первым и вторым лицом, в значительной степени в дополнительной дистрибуции, наблюдаются достаточно сильные семантические различия. В данной работе мы демонстрируем специфичность коллокаций глаголов мышления и глаголов физического действия и обсуждаем значение лексико-синтаксической идиосинкразии для когнитивной лингвистики.

Ключевые слова: мышление, корпусные методы, коллокация, лексико-синтаксический анализ, инфлексивные категории, эпистемизация

1. INTRODUCTION

In the present study, we describe a number of relatively low-level patterns associated with basic verbs of ideation (THINK, KNOW), along with other peculiar inflectional patterns in a miscellany of constructions¹. In the case of the ideation verbs, it is the specific combination of subject pronouns and these verbs that will be our focus. Our interest lies in identifying recurring patterns of usage and, where possible, seeking motivation for such patterns in human experiential realities. This approach to the study of language, grounding language phenomena in broader cognitive realities, is rightly called a cognitive linguistic approach (see Dancygier 2017a for an introduction to the field of cognitive linguistics as currently practised and Dancygier 2017b for contemporary overviews of subfields). Our adoption of a corpus-based methodology to investigate these patterns reflects, too, a widely held view within cognitive linguistics that a usage-based approach is a tool of critical importance. Indeed Dancygier (2017a: 2) remarks that “actual usage is at the core of cognitive linguistics”.

Our decision to focus on the pair {THINK, KNOW} is based on a number of considerations: the relatively high frequency of such verbs in ordinary discourse; the closeness of each member of the pair to the other semantically, creating potentially interesting contrasts in the details of usage; the tendency for each of these verbs to become discourse markers. These considerations suggest that these two verbs have affinities with each other that can be profitably studied at a finer-grained level than has been done to date, giving insight into why they each take different paths in terms of semantic shift and why they manifest highly skewed and individualized agreement patterns.

More specifically, the aims of this study are (i) to identify statistically significant combinations of subject/object pronouns with select English verbal expressions using corpus-based methodologies; (ii) to identify preferences for the use of 1st person in other miscellaneous constructions, prompted by our findings from (i); and (iii) to reflect on the larger significance of our findings for the field of cognitive linguistics. We begin with some relevant background research on co-occurrence patterns of number/person and verb categories (§2). We then introduce the corpus and the statistical methods used in the study (§3), present the findings (§4), and discuss the larger significance of the findings for cognitive linguistics (§5).

¹ We use small caps (e.g. THINK) for any lemma (and for referring to lemma equivalents across languages) and italics (e.g. *think, thinks, thought*) for all inflected forms.

2. BACKGROUND

The co-occurrence patterns of pronouns with certain verbs have already received attention in the linguistics literature as part of the typological interest in the Hale-Silverstein person hierarchy of 1st > 2nd > 3rd (Hale 1972, Silverstein 1976), but there has been rather less interest in patterns occurring with specific *inflected* forms of verbs. For English, statements about co-occurring argument types (whether it is the semantics of the arguments or how hierarchies of person or animacy play out) are usually made at the lemma level.

In the context of corpus linguistic research, Sinclair (1991:8) suggested that inflectional differences may be more important in terms of their patterning than is commonly assumed, taking the inflected word form (rather than a lemma) to be the default unit of study: “There is a good case for arguing that each distinct form is potentially a unique lexical unit, and that forms should only be conflated into lemmas when their environments show a certain amount and type of similarity.” For example, in Sinclair’s illustration of this approach, adjectival forms like *bloody* and *bloodiest* are kept apart in a word count of a corpus, as are *is* and *are*. Sinclair’s position has been recently restated by Knowles & Don (2004: 71): “...it has become apparent that individual members of the lemma can behave independently and develop their own meanings and collocations”². Newman & Rice’s (2006: 31) notion of an “inflectional island” is very much in the same vein as Knowles and Don’s remarks, referring to syntactic/semantic properties that tend to inhere in individual inflections of a verb, rather than extending across all inflected forms of the lemma. In that paper, Newman and Rice found distinctive and intriguing patterns of PRO subjects with transitive and intransitive uses of EAT and DRINK verbs in spoken and written registers.

Recent research into patterning at the inflectional level has yielded promising results (cf. studies exploring quite specific lexical items such as Thompson & Mulac 1991, Aijmer 1997, Kärkkäinen 2003, and Van Bogaert 2011 on *I think*; Tao 2001, 2003 on *remember* and *forget*). Many of these studies focus on the grammaticalization of what have been termed complement-taking mental predicates into complement-less pragmatic markers. Unlike our study here, the majority of these previous analyses have not been corpus-based, although they have appealed to familiar corpus notions such as high frequency and increased collocational fixedness that do have a bearing on grammatical entrenchment or what Schoonjans (2012) has called “particulization”. He uses this concept in the context of the German ideational verb *glauben* ‘think/believe’, which has both lost its 1sg pronoun, *ich*, and its TAM (tense-aspect-mode) inflection and emerged as a sentence-medial modal particle, *glaub*, with the evidential force of ‘maybe’ or ‘perhaps’. While the present study is consonant with much of that prior grammaticalization research, our purpose is to examine why such grammaticalization came about in the first place through heavily skewed inflectional preferences (for 1sg.pres and 2.pres, respectively) affecting the major ideational verbs. Our aim is not to relitigate the case for the grammaticalization of these verbs into pragmatic markers, but to show how

² See also the references to earlier studies on person and number preferences with verbs in Scheibman (2001: 61—63, 2002: 1—87).

first-person singular (*I*) and second-person (*you*) ideation are associated with different semantic values which have had huge consequences for the incipient epistemization of non-ideational verbs. Not only do different types of predicates enter into the ideational arena, but they tend towards specific inflectional and collocational preferences, as we will show through a series of corpus searches and analyses. In short, the ways that “I ideate” as opposed to “you ideate” are strongly linked to connotations of *I think* and *you know* in the first place. These two epistemic constructions are different and differentially draw non-epistemic verbs and constructions into their respective orbits or, as we describe in §5, their respective “attractor basins” (in the sense of Spivey 2008).

Another uniquely valuable contribution to our understanding of inflectional level patterning in English is Scheibman’s (2001) discussion of subject types, sub-categorized in terms of person and number, with different classes of verbs. She throws light on the notion of “subjectivity”, understood as how speakers and writers use linguistic devices to express their own individual perceptions, feelings, and opinions. As in our present study, Scheibman’s (2001) research focuses on the preference for certain person and number choices (1st person singular, 2nd person singular, etc.) as grammatical subjects of verb types and her verb classes include cognition (*know, think, remember, figure out, etc*) verbs. While her study of these larger classes (alongside other broad categories) is helpful, especially when it comes to comparing results across lexical fields, we have chosen to explore linguistic patterning at a more fine-grained level, reporting on patterns involving selected individual verbs and expressions, i.e., *think vs. know, go vs. run through one’s mind, etc.*

It is appropriate to mention, too, relevant research in Natural Semantic Metatheory (NSM; cf. Wierzbicka 1996, Goddard 1997, Goddard & Wierzbicka 2014). While NSM does not employ the highly quantitative methods of some of the works mentioned above, it succeeds in providing insightful semantic analyses building upon a set of semantic primitives. It is not a coincidence that the verbs we have chosen to focus on are among the six mental predicates recognized in later versions of the inventory of semantic primitives in NSM, namely THINK, KNOW, WANT, FEEL, SEE, HEAR (using small caps here to denote these primitives, following the practice in NSM). NSM shares the broader cognitive linguistic interest in the role of ordinary bodily realities and experiences in motivating and shaping aspects of language behavior and it is not surprising that our own approach has brought us to a set of verbs that play a key role in NSM. The discussion of *I think* and miscellaneous other epistemic phrases of English in Wierzbicka (2006: Chapter 7) shows a further overlap between NSM and our own focus in this study. It is of interest to note that when they occur in definitions of words, these mental predicates in NSM may sometimes appear specifically with the 1SG pronoun. So, for example, the 1SG pronoun is required as the subject of WANT in the sequence of statements “many good things are happening to me now as I want; I can do many things now as I want; this is good” as part of the explication of *He was happy* (Goddard and Wierzbicka 2014: 103). In other publications, too, Wierzbicka has turned attention to the different semantic content associated with different choices of number/person subjects in expressions, e.g. 1SG and 3PL (e.g. *people*) subject frames of *to have a sense that* (Wierzbicka 2010: 169—176).

3. METHODOLOGY

Throughout this study, we rely upon the Corpus of Contemporary American English (COCA, <https://corpus.byu.edu/coca/>) for our English usage data. COCA is a corpus of contemporary American English (Davies 2008-) and has been tagged using the CLAWS 7 tagset. It is available to users via a web interface, which is how it was accessed for this study. The corpus consists of texts dating from 1990—2017 and is being added to each year (thus, it is a “monitor corpus”). We see spoken language as being particularly relevant in the present study, since it is in spoken language that one might expect to see a greater representation of emergent constructions. Our corpus searches will therefore be restricted to the spoken component of COCA or what we will call COCA_{sp}. COCA_{sp} consists of transcripts of unscripted conversation from more than 150 different TV and radio programs, making up over 118 million words at the time of writing (2018). The programs on which COCA_{sp} is based are largely concerned with American news and current affairs, along with some idiosyncratic interview-style programs. As such, the language of COCA_{sp} may be called naturalistic for these contexts because it is interactional, but it is not necessarily natural as far as ordinary, everyday conversation is concerned.

Regardless of grammatical case, we will refer to the six pronoun forms under investigation (1SG, 2, 3SGM, 3SGF, 1PL, 3PL) simply as the *pronouns* (PRO) without further qualification. Both upper-case and lower-case forms of the pronouns will be included in frequency counts. The decision to exclude *it* relates to specific interest in verbs of sentience occurring with animate, especially human, participants, rather than with inanimate. Sequences such as [PRO + verb] will be used as the basic proxy pattern for retrieving personal subject pronouns occurring with the verb forms. A refinement of this search pattern may be used to find the [subject PRO + present tense verb] sequences in declarative structures such as *such She knows a lot* and *What you know about dinosaurs is amazing*, but not *Does he know anything about dinosaurs?* While interrogative structures would be a viable and interesting extension of the present study, they will not be included here. In terms of “precision” (i.e., how well our returns match subject and verb combinations), our proxy search for PRO as the subject of a verb is high, attributable in part to the availability of the CLAWS 7 part of speech tags on the verbs, distinguishing present tense forms (vv0, vvz) from infinitival forms (vvi). The “recall” (i.e., the extent to which our returns include all the relevant subject-verb combinations), on the other hand, is not 100%. Subjects of verbs are, of course, not restricted to the position immediately to the left of the verb, even in declarative structures; rather, they can appear some distance to the left. With pronominal subjects, there is less likelihood of intervening relative clauses than with nouns (as in *He who thinks before acting is wiser*), but certainly adverbials can easily intervene (as in *I always think of her*). Recall is clearly not ideal, but, importantly, we use the same kind of search pattern in each case and the comparison across the search results is based on the methodological decision to use the same position immediately to the left of the verb in most searches. (We make an exception in the case of certain adverbs like *suddenly*, discussed below in §4.2.2.)

While frequency of occurrence of patterns lies at the heart of this study, we will make use of a more sophisticated (but easily understood) statistical measure in reporting on the verbs that are the main focus, i.e., the epistemic verbs, for which we have sufficient frequencies to test statistically. The statistical measure involves a calculation of *standardized residuals* associated with a chisquare statistic, indicating the extent to which particular pronouns occurring as the subject/object of a verb are overused or underused. In considering the patterning of PRO as the subject of a verb, the initial step is to determine the frequencies of the combination [PRO + present tense of any verb] in COCA_{sp}³. These frequencies may be called the baseline frequencies and are shown in Table 1. From these frequencies, we can see the proportions of *I*, *you*, *(s)he*, etc. in the whole corpus functioning as the subject of a verb in the present tense, expressed as percentages in Table 1.

Table 1

**Baseline frequencies and percentages of occurrence of subject pronouns
of all lexical verbs (base or present-tense forms) in COCA_{sp}**

CLAWS Tag	PRO _{subj}	Raw Frequency	Row Total / Column Total as Percentage	Rank Order
[vv0]	<i>I</i>	718,184	39%	1
[vv0]	<i>you</i>	599,585	33%	2
[vvz]	<i>(s)he</i>	145,014	8%	5
[vv0]	<i>we</i>	207,887	11%	3
[vv0]	<i>they</i>	163,885	9%	4
Total		1,834,555	100%	

The frequencies of the pronouns occurring with the present tense of any verb in COCA_{sp} are “expected” to be in the same proportions as the overall proportions in Table 1 (or, more weakly, to share the same overall rank order). That is, we start with the assumption that the proportion of some phenomenon in a sub-part of the population will be identical to that found in the whole population (the “null hypothesis”, cf. Gries 2013b: 316—319) and proceed to show how likely this assumption is given the discrepancies between the observed and expected frequencies of the skewed agreement phenomenon we are investigating. Once the expected frequencies have been calculated, it is possible to compare them with the observed frequencies and evaluate the statistical significance of the difference between them, as in a chisquare test. The *standardized residuals* represent a standardized value of the difference between the observed and expected frequencies for each combination of pronoun and verb implemented in R (R Development Core Team 2014), obtained by calculating the differences between observed and expected frequencies, divided by the square root of the expected frequency (Agresti 2007: 38—39). Standardized residuals with values greater than +2 or less than

³ By “any verb”, we mean any “lexical verb”. This category excludes forms of DO, HAVE, BE, all of which have special properties in their auxiliary verb use. The frequencies in Table 1 were obtained by a series of searches using the strings “I.[pp*] [vv0]”, “you.[pp*] [vv0]”, “she.[pp*] [vvz]”, “he.[pp*] [vvz]” etc., with the “Chart” option selected to display total frequencies for each genre (“Section”) in COCA. All corpus findings were the result of searches in COCA during February 2018.

–2 indicate statistically significant overuse or underuse in those cells. It is also helpful to display the overuse and underuse of pronouns with verb forms graphically, as in an association plot (cf. Gries 2013a: 187—188), and we will make use of these plots in the course of our exposition⁴.

4. PRONOUNS + VERB PREFERENCES

4.1. The ideational verbs, THINK and KNOW

We begin our discussion with the distinct frequency profiles of inflected forms of THINK and KNOW in spoken English (here, in their simple present tense forms) with different agreement patterns as measured by their co-occurrence with the different subject personal pronouns. As we will argue in §5, these distributional differences have had a concomitant effect on the recruitment of non-ideational predicates and constructions to take on epistemic meanings in the language. We queried COCA_{sp} for all subject pronouns (except *it*) occurring with base or present-tense forms of *think* and *know*, using the POS (part of speech) tags, vv0, vvz, and compared those frequencies with all other lexical verbs in the spoken sub-corpus occurring with the same set of pronouns. Table 2a gives the raw (observed) frequencies for *think(s)* with a pronominal subject while Table 2b gives the standardized residuals when compared with the raw frequencies for all other verbs.

Table 2

(a) Observed frequencies for THINK [PRO + *think.vv0* | *vvz*] compared to frequencies of all other lexical verbs (base or 3SG present-tense forms) in the spoken subcorpus of COCA.

(b) Standardized residuals for THINK frequencies compared to frequencies of all other lexical verbs

(a)	<i>I</i>	<i>you</i>	<i>(s)he</i>	<i>we</i>	<i>they</i>
THINK	254,252	15,855	828	6,767	5,996
other	463,932	583,730	144,186	201,120	157,889
Total	718,184	599,585	145,014	207,887	163,885

(b)	<i>I</i>	<i>you</i>	<i>(s)he</i>	<i>we</i>	<i>they</i>
THINK	599.0677	-334.6267	-163.4517	-163.5032	-138.5127
other	-599.0677	334.6267	163.4517	163.5032	138.5127

Figure 1 shows the association plot for the distribution of *pro* x *think*⁵. For present purposes, it is the relative height of the boxes, reflecting the values of the standardized residuals, that is most relevant. The black rectangles in Figure 1 show the overuse of subject pronouns with present-tense forms of verbs in a more immediate and more striking way than by inspecting numerical tables. 1sg, while hugely overrepresented with THINK, is greatly underrepresented across the rest of the verbal lexicon, on average.

⁴ We used the functions `chisq.test()` and `$std` for the standardized residuals and `assocplot()` to create the association plots in the base package of R (R Development Core Team, 2018).

⁵ In the association plots shown here, black rectangles above a dashed line correspond to cases where the observed frequency is greater than the expected frequency; white rectangles below a dashed line correspond to cases where the observed frequency is less than the expected frequency. The area of a box in such plots is proportional to the difference in observed and expected frequencies; the width of the rectangle is proportional to the square root of the expected frequency; the height of the rectangle is proportional to the standardized residual.

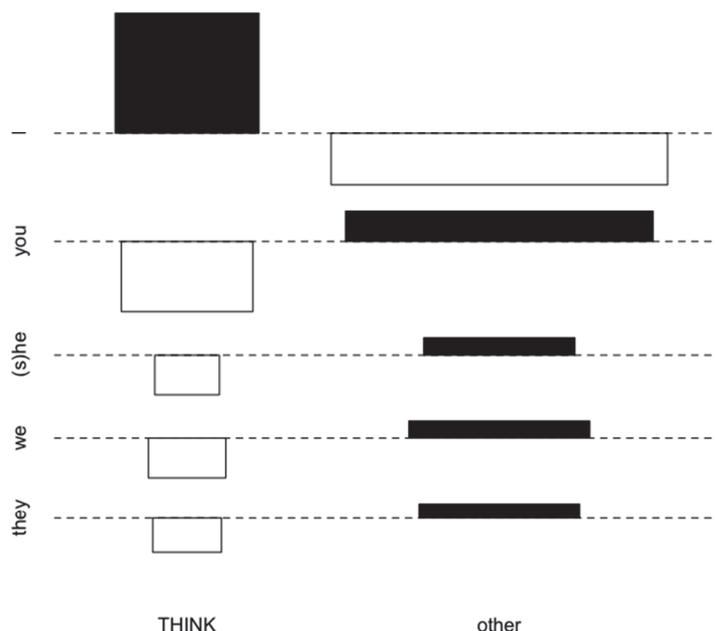


Figure 1. Association plot of PRO x THINK compared to other verbs in COCA_{sp}

The PRO x THINK distributions compared to PRO x OTHER VERBS distributions for THINK are but half of the story. When we look at the distributional frequencies for KNOW, we start to get a picture of the differential behaviour and distinct epistemizational attraction to concepts of first person singular ideation versus second person ideation. Table 3 gives the raw (observed) frequencies for *know(s)* with a pronominal subject as well as the standardized residuals when compared with the raw frequencies for all other verbs. Figure 2 shows the association plot corresponding to the distribution of pronouns given in Table 3. Comparing Figures 1 and 2, we see how *know* is the converse of THINK.

Table 3

(a) Observed frequencies for KNOW [PRO + *know.vv0* | *vvz*] compared to frequencies of all other lexical verbs (base or present-tense forms) in COCA_{sp}.

(b) Standardized residuals for KNOW frequencies compared to frequencies of all other lexical verbs

(a)	<i>I</i>	<i>you</i>	<i>(s)he</i>	<i>we</i>	<i>they</i>
KNOW	42,219	320,202	16,809	5,891	2,815
other	675,965	279,383	128,205	201,996	161,070
Total	718,184	599,585	145,014	207,887	163,885

(b)	<i>I</i>	<i>you</i>	<i>(s)he</i>	<i>we</i>	<i>they</i>
KNOW	-406.1812	745.5418	-92.84945	-217.1432	-201.8366
other	406.1812	-745.5418	92.84945	217.1432	201.8366

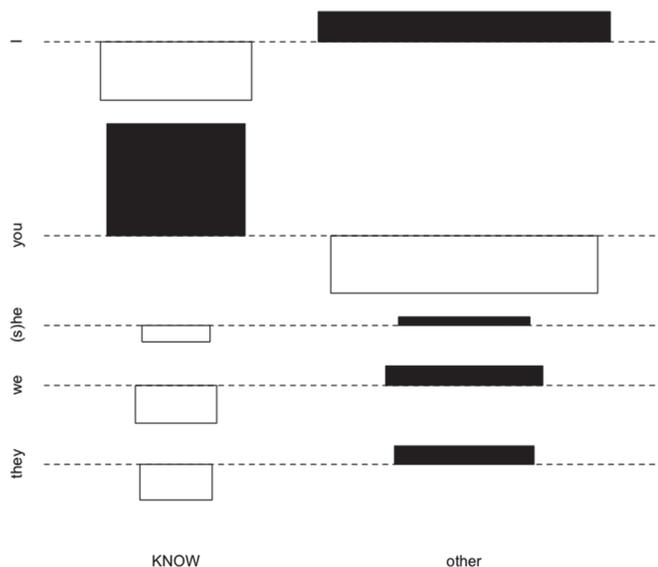


Figure 2. Association plot of PRO x KNOW compared to other verbs in COCA_{sp}

Our main purpose in this section is to establish, statistically, the attraction that THINK and KNOW have for particular inflectional forms of the subject pronouns, especially 1sg and 2 person subjects, rather than explore the particular constructions in which these combinations occur. The syntactic, semantic, and pragmatics of the individual uses of these verbs is beyond the scope of the present study. Even so, it is of interest to note the co-occurrence, indeed the juxtaposition, of *I think* and *you know* in examples such as (1a-b). In these examples, illustrating the two possible orders *I think* + *you know* and *you know* + *I think*, we see *think* used with a clause complement while *you know* appears as a complement-less pragmatic marker (cf. §2).

- (1) a. ***I think, you know, everything changes in politics, but I've, I've...*** (SPOK: NBC_MeetPress, 2007)
 b. ***You know, I think we will see that eventually.*** (SPOK: NPR_ATCW, 2006)

Although we are relying on the standardized residuals to establish the statistical significance of the overrepresentation of *I think* and *you know* in the corpus, it is still instructive to consider some relevant raw frequencies related to the use of subject pronouns and lexical verbs in the present tense in the corpus. Table 4 lists the 20 most frequent base forms ([vv0]) occurring immediately to the right of *I* and *you*, respectively. In this table, we can readily see the overall preference for verbs of cognition (*think, mean, know, want, guess, remember, understand, etc.*) in this construction with *I*. THINK is not just the top-ranked verb in the first column of this table, it enjoys nearly two and a half times the frequency of the second-ranked verb, MEAN (707,880 vs. 284,116). The fourth column lists the results for the 20 most frequent base forms co-occurring with *you*. One sees in these results a greater variation in the semantics than with the *I*-verbs, with non-cognition verbs such as GO, LOOK, SAY, TALK, COME, FIND, etc. making a conspicuous appearance in the list. Here, know is far and away the most

frequent verb (320,202), well ahead (at 12 times the frequency!) of the second-ranked verb, WANT (26,762). In other words, the particular preferences for *I think* and *you know* that we see in Figures 1 and 2 do not tell the whole story about the attraction of these verbs to 1sg and 2nd person subjects, respectively; these preferences are evident in a striking way even when all present tense verbs are considered. Nevertheless, as the two most frequent [PRO-verb.PRES] bigrams in COCA_{sp}, we have to acknowledge that *I think* and *you know* are, individually, huge constructional magnets for other expressions. It is incumbent, then, that we come to understand the particular semantic associations and connotations that imbue *I think* and *you know* since we find equally skewed distributions (by subject, object, or possessive pronoun agreement) with non-ideational expressions that have come to have epistemic force in the language, even though they were originally verbs of perception or physical action. Happily, corpus analysis can help us do this.

Table 4

The 20 most frequent sequences of *I* vs. *you*, respectively, with present tense lexical verbs (queried as [vv0]) in COCA_{sp}

1SG + verb.PRES	frequency	rank	2 + verb.PRES	frequency	rank
<i>I think</i>	707,880	1	<i>you know</i>	320,202	1
<i>I mean</i>	284,116	2	<i>you want</i>	26,762	2
<i>I know</i>	201,758	3	<i>you get</i>	17,431	3
<i>I want</i>	167,388	4	<i>you see</i>	16,810	4
<i>I guess</i>	82,616	5	<i>you think</i>	15,855	5
<i>I love</i>	68,174	6	<i>you look</i>	15,836	6
<i>I believe</i>	67,618	7	<i>you go</i>	14,708	7
<i>I feel</i>	62,274	8	<i>you say</i>	13,350	8
<i>I like</i>	54,782	9	<i>you need</i>	10,109	9
<i>I see</i>	54,498	10	<i>you take</i>	4,743	10
<i>I say</i>	53,060	11	<i>you talk</i>	4,631	11
<i>I hope</i>	49,798	12	<i>you put</i>	4,499	12
<i>I remember</i>	44,954	13	<i>you hear</i>	4,264	13
<i>I get</i>	41,936	14	<i>you like</i>	3,818	14
<i>I need</i>	39,570	15	<i>you feel</i>	3,808	15
<i>I understand</i>	30,592	16	<i>you make</i>	3,449	16
<i>I wish</i>	27,346	17	<i>you mean</i>	3,380	17
<i>I go</i>	23,902	18	<i>you start</i>	2,812	18
<i>I look</i>	23,670	19	<i>you come</i>	2,770	19
<i>I suppose</i>	22,622	20	<i>you find</i>	2,199	20

4.2. Miscellaneous activity verbs and constructions

Having established (i) that the basic verbs THINK and KNOW have highly skewed inflectional profiles and (ii) that *I think* and *you know* are uniquely privileged uses of THINK and KNOW and wildly dominant inflectionally speaking, we now turn to a range of verbal constructions that, when taken literally, have nothing to do with ideation, but which clearly have undergone epistemization processes in the language. That is, certain verbal constructions are turning into expressions about ideation or knowledge validation and they are turning up with highly skewed inflectional profiles of their own.

While we find these expressions interesting as cognitive linguists because they have taken on meanings beyond the literal and the physical, they prove to be especially fascinating to us as corpus linguists because they display similar inflectional skewing as we find with the two basic cognition verbs explored in §4.1. Moreover, through an examination of their frequencies by agreement and TAM, we can gain insight into how first person singular ideation is construed in English, compared to ideation affecting second persons. If language change or semantic shift is driven in part by analogy, then a better understanding of the different semantic associations affecting verbal expressions by person helps us make the larger point advocated by Sinclair and others that the inflected lexical form is the proper starting point for lexico-syntactic analysis, not the idealized lemmatized form. In this section, because the frequency counts are relatively low, we will only report raw frequency with no further statistical analysis. It is worth noting that the cognizer in the following constructions are not encoded as the subject of the verb, but as a down-stream thematic participant, construed as a patient or as the object of a preposition. The subject is generally a pleonastic, *it*, or the headless relative pronoun, *what*.

4.2.1. *It/What* STRIKE/HIT PRO

The two physical verbs that have re-lexicalized or, actually, constructionalized into verbs of ideation the most are STRIKE and HIT. Indeed, the participial adjective, *striking*, collocates most frequently with nouns that are associated with epistemic realization or discernment, such as *thing, resemblance, contrast, example, difference, and similarity*. With STRIKE and HIT, the cognizer presents as the direct object, as in *it struck me* or *what hit him* or as the prepositional object with progressive forms of strike, as in *it was striking to me*; therefore, our COCA_{sp} searches involved variants of these search strings: [*what|it* [strike|hit] (p*)] or [*what [BE] striking* (to) (p*)]. Tables 5 and 6 summarize the returns for STRIKE and HIT, respectively, by TAM and construction (cognizer is a pronominal DO or X). Corpus examples follow in (2) and (3).

Table 5

**Observed frequencies for [*what/it* STRIKE (to) PRO] in COCA_{sp}.
[*All 44 instances of *what struck you* are questions, with the *what* functioning as a bona fide question word, as opposed to the function of *what* to introduce a pseudo-cleft as in (2a).]**

strike	<i>me</i>	<i>you</i>	<i>him</i>	<i>her</i>	<i>us</i>	<i>them</i>	Row Total
<i>what strikes PRO</i>	111	26					137
<i>what struck PRO</i>	144	44*	1	1	4		194
<i>what's striking PRO</i>	1						1
<i>what is striking to PRO</i>	3						3
<i>what was striking to PRO</i>	6						6
<i>it strikes PRO</i>	191	3	2		3		199
<i>it struck PRO</i>	105	3	2	2	6	3	121
<i>it is striking to PRO</i>	3						3
<i>it was striking to PRO</i>	2						2
<i>Column Total</i>	566	76	5	3	13	3	666
<i>Column Total / Row Total as Percentage</i>	85%	11%	< 1%	< 1/2%	2%	< 1/2%	100%

- (2) a. *He can be personable, but he also can be very serious. Now **what is striking to me** is that he still seems removed from who he’s working for. I mean, he was clearly referencing Harvey Weinstein in respect for women, or that would be the best guess* (SPOK: CNN_Anderson Cooper, 2017).
- b. *It’s odd that he—well, **it strikes me** as a little bit odd that he continually talks about his struggle to get there, the reasons behind it, struggle, it’s well established, it’s done. I have a feeling he should stop talking like that if there’s any possibility that he’s going to get in this thing...* (SPOK: CBS_FaceTheNation, 2015).
- c. *Before this kind of gradual, almost indistinguishable, process of- of corruption. You know, **it struck me again**, in- in- in going back through it, how much it really is a parable of the dark side of the moon of the democratic proposition...* (SPOK: NPR_Weekend, 1996).

Table 6

Observed frequencies for [what/it HIT PRO] in COCA_{sp}. Note the pseudo-cleft uses of *what* in (3a-b). The relatively high values for *it hit/hits you* shouldn’t be taken at face value. Nearly all of them are used generically or refer back to the 1sg speaker, as in (3d)

hit	me	you	him	her	us	them	Row Total
<i>what hits PRO</i>	2	1				1	4
<i>what hit PRO</i>	9	7	12	1	3	8	40
<i>what’s hitting PRO</i>						1	1
<i>it hits PRO</i>	13	20	3		1	4	41
<i>it hit PRO</i>	79	16	4	6	4	2	111
<i>it’s hitting PRO</i>	1		2		1	1	5
<i>it is hitting PRO</i>	1						1
Column Total	105	44	21	7	9	17	203
Column Total / Row Total as Percentage	52%	22%	10%	3.5%	4.5%	8%	100%

- (3) a. *Well, I think that-**what hits me** about this ethics stuff, Robert, is that I’m surprised that the Democrats don’t see an opening with campaign finance reform.* (SPOK: NPR_ATC, 1995)
- b. *One of the lawyers for detainees approached me and said, I want my clients’ art to be exhibited. I said, what do you mean? There’s art made at Guantanamo. **What hit me** at first was how normal they seem. Shouldn’t their drawings be so much more angry?* (SPOK: PBS_Newshour, 2017)
- c. *Every once in a while **it hits me**. I’ll be driving along or whatever, just by myself, and start to think about it, and it was really close. I mean any — a foot either way, an inch either way, it would have been over for me.* (SPOK: ABC_PrimeTime, 1994)
- d. *You know, after the firefight’s over and the adrenaline rush is over and you started — you know, you’re all soaking wet and just feel like your legs won’t hold you, you know, **it hits you**. I just took a life.* (SPOK: Dateline_NBC, 2008)

It is apparent from the counts, the examples, and the brief commentary in this section that STRIKE and HIT, when used to convey mental (not physical) force, have an overwhelming preference for 1sg objects construed as the target of sudden realization. More than three-quarters of the returns in COCA_{sp} for this family of [what/it STRIKE/HIT PRO] constructions are about 1SG ideation. Obviously, STRIKE and HIT bring many

semantic associations from the physical world when used figuratively. They both suggest punctual, telic, and dynamic action, which, we argue, carries over into how 1SG ideation is construed more generally. We return to this point in §5.

4.2.2. It DAWN on PRO

For *it to dawn on someone* is a particularly nice figurative expression in English to describe epistemic realization. The various TAM-inflected forms of what is otherwise a concrete verb, DAWN, describing the path of the sun and the return of daylight (widely associated with consciousness and understanding), show an overwhelming preference for 1SG prepositional objects, the nominal relation that encodes the cognizer in this construction. Table 7 shows the raw frequencies from COCA_{sp} by TAM and person of the prepositional object. We have broadened the searches to also include the adverbs which collocate with *it dawns/dawned on*. We provide some actual corpus returns in (4).

Table 7

Observed frequencies for [*it* (ADV) DAWN on PRO] in COCA_{sp}

dawn on	<i>me</i>	<i>you</i>	<i>him</i>	<i>her</i>	<i>us</i>	<i>them</i>	Row Total
<i>it dawns on PRO</i>	3			1		1	5
<i>it dawned on PRO</i>	33	3	9	1	1	2	49
<i>it never dawned on PRO</i>	9				2		11
<i>it just dawned on PRO</i>	8			1			9
<i>it really dawned on PRO</i>	8	1					9
<i>it finally dawned on PRO</i>	4		3		1		8
<i>it suddenly dawned on PRO</i>	6					1	7
<i>it slowly dawned on PRO</i>	1				2		3
<i>it gradually dawned on PRO</i>	1				1		2
<i>it fully dawned on PRO</i>	1						1
<i>it probably dawned on PRO</i>			1				1
Column Total	74	4	13	3	7	4	105
Column Total / Row Total as Percentage	71%	4%	12%	3%	6%	4%	100%

- (4) a. *I was meeting people and talking to them, and while talking to them, it dawned on me — oh, my God, Nick, you're talking about something I've heard of. I know this song.* (SPOK: NPR_Fresh Air, 2015)
 b. *I don't know how we were lucky enough to figure that out, but — and it suddenly dawned on me what I truly have in common with everybody else is this one man's vision affected all of us.* (SPOK: NPR_Sunday, 2000)

A brief final point about the [*it DAWN on PRO*] construction and the collocating adverbs listed in Table 7. For the most part, the rather absolute and categorical *never*, *just*, *finally*, and *fully*, along with the intensified *really* and *suddenly* show a marked preference for 1sg, as does the construction as a whole. We do not regard it as incidental that the less forceful or dynamic adverbs *slowly* and *gradually*, or the indeterminate *probably*, show a slight preference for non-1sg cognizers. A point we make in §5 is that a range of somewhat covert semantic notions like these seem to be attached to the way 1st person singular ideation is construed. These are not necessarily associations evident in *I think*, but which nevertheless guide the non-ideational expressions that come to take on epistemic force towards or away from 1SG.

4.2.3. V_{stasis} in/on PRO's mind

A thought, idea, or bit of knowledge can be *in* or *on one's mind* in English. Such expressions suggest simple, stative locative constructions, far from the dynamic construals afforded by the likes of *it struck someone*, *it hit someone*, or *it dawned on someone* examined previously. Nevertheless, the fact that these expressions are based on a spatial metaphor and metonymy [viz. the place (*mind*) is the locus of activity (*thinking*) happening in that place]. While 1SG cognizers were prevalent in those other constructions, the more static, *in/on one's mind*, show only modest preferences for 1st person. Table 8 presents the returns from COCA_{sp} for *in PRO's mind* and *on PRO's mind*, respectively. A handful of actual returns from the corpus follow in (5).

Fewer than half of the examples in COCA_{sp} of the [*in/on PRO mind*] construction involve a 1SG cognizer (in the form of the possessor of *mind*). Indeed, these constructions seem to be better distributed across all the potential sentient players: 1SG (47%), 2 (30%), 3SGM/F (17%), 1PL (3%), 3PL (3%), in proportions far closer to those “background” frequency distributions reported in Table 1 for all lexical verbs, as represented in COCA_{sp}. The rank order is nearly the same, for example: 1SG (#1), 2 (#2), and then the rest at a distance. The lack of overwhelming attraction to 1SG suggests that the semantic properties associated with the fairly stative and locative [*in/on PRO mind*] construction are fairly neutral, person-wise. As we'll see below, the more active and dynamic the figurative expression, the more it displays an attraction to a 1SG cognizer.

Table 8

Observed frequencies for [*in*] on [app*] *mind*] in COCA_{sp}

	<i>my</i>	<i>your</i>	<i>his</i>	<i>her</i>	<i>our</i>	<i>their</i>	Row Total
<i>in PRO's mind</i>	1,915	1,105	384	128	56	134	3,722
<i>on PRO's mind</i>	205	235	189	55	74	19	777
Column Total	2,120	1,340	573	183	130	153	4,499
Column Total / Row Total as Percentage	47%	30%	13%	4%	3%	3%	100%

- (5) a. *I mean — I thought and, you know, knowing how I felt about him. I was angry, because **in my mind** he was doing that to—that was like his parting gift, right?* (SPOK: CNN_The Lead with Jake Tapper, 2017)
- b. *You know, I found that some of them never even pulled a gun out. They shoot — you know, they just reached down and grabbed the gun and twisted their holster and fired right through the holster. So **in your mind**, you think because we've always shown Westerns that they take it out and shoot — some of them never took them out.* (SPOK: NPR_Fresh Air, 2016)
- c. *you don't wake up in the morning and immediately start thinking about that. What's on people's minds is what's **on your mind** and my mind and everybody else's mind, and that is how am I going to provide for my family?* (SPOK: CBS_ThisMorning, 2012)
- d. *That's why it's weighing very heavy. It's been weighing heavy for 37 years **on his mind**. I think he really wants to tell it.* (SPOK: NBC_Dateline, 2005)

4.2.4. V_{motion} through PRO's mind

Similar to the *in/on one's mind* expressions just examined, an idea, thought, or realization can pass *through one's mind*, in a slightly more dynamic fashion. Because motion verbs are involved, we have naturally categorized these as activity expressions. Admittedly, the epistemic or ideational sense is brought about figuratively by the presence of the locative nominal, *mind*. However, the choice of verb is somewhat affected by the choice of possessive pronoun in ways reminiscent of the 1SG vs. 2 person differences noted above in other expressions. Far and away, the most frequent verb to enter into this construction is *go*, a nearly manner-less verb of motion. The more force-dynamic the verb, however, the more likely it is being used to express ideation in *my mind/head*. Table 9 gives the lemmatized frequencies for ideation constructed with verbs of motion through the mind or head. Some actual returns from COCA_{sp} are presented in (6).

Table 9

Observed frequencies for [vv0/vvz through [app*] mind] in COCA_{sp}. Here, we only present lemmatized results because we are specifically focusing on choice of verb with different prepositional objects, not TAM forms. The verbs are impressionistically arranged in order of increasing degree of energy, punctuality, or forcefulness, rather than frequency. When the locus of motion is *head*, not *mind*, the counts are given in parentheses. Aggregate counts are given in the row marked Total

VERB _{motion} through PRO's mind (head)	my	your	his	her	our	their	Row Total
<i>float</i>		1					1
<i>roll</i>	1	(1)					1 (1)
<i>slide</i>	1						1
<i>pass</i>	2	2					4
<i>cross</i>	2	1					3
<i>come</i>	2	1					3
<i>go</i>	130 (43)	490 (84)	36 (10)	22 (9)		4 (1)	682 (147)
<i>run</i>	19 (20)	10 (3)	7	1 (1)		1	38 (24)
<i>rush</i>	1 (1)	1					2 (1)
<i>race</i>	3	(1)					3 (1)
<i>jump</i>	1						1
<i>flash</i>	8 (3)	(1)					8 (4)
Column Total	237	596	53	33		6	925
Column Total / Row Total as Percentage	25%	64%	7%	3.5%		.5%	100%

- (6) a. Give me more of what **went through your mind** as you read this thing, that made you say, "I have to have it". (SPOK: CBS_SundayMorning, 1993)
 b. And when [she] said to me, "Well, how are you doing?" I, you know, it just **flashed through my mind**. I have asked that question to people in so many settings across our country and even the world for so many years, and I felt like, you know, we're finally understanding that that's what it's about. (SPOK: CBS_GMA, 2008)
 c. And, Terry, you've been listening to this as well. I'm just wondering if any other thoughts **floating through your mind** about this? (SPOK: CNN_TalkBack, 2002)

Further to the discussion in §4.2.3, we see a definite preference for 2nd person cognizers in figurative expressions suggesting that ideation involves movement (of a concept or percept) through one’s mind or head. That said, we add the proviso that this agreement preference holds only for the most manner-less motion verb, GO, which nevertheless accounts for nearly 90% of the examples. With motion verbs that conflate manner or path or, especially, an active dynamism, as in RUN, RACE, or FLASH, the agreement preference tilts back to 1SG. Of note is the fact that the one instance of the rather passive, [FLOAT *through one’s mind*], involves a 2nd person cognizer.

4.2.5. “lightbulbs” and other “suddenly realize” expressions

Rounding out this discussion of miscellaneous, physical domain expressions that have taken on an ideational or epistemic reading are a pair of constructions that revolve around concrete nouns, specifically *lightbulb* and *penny*. In English, one can say that *a lightbulb went off/on in/over one’s head*, meaning that one has suddenly had a realization about something. This imagery-rich concept for realization only seems to be attested since the 1960s on Google N-gram Viewer (<https://books.google.com/ngrams/>) and only three examples are available from COHA, the Corpus of Historical American English, a sister corpus to COCA (available at <https://corpus.byu.edu/coha/>), all three of which are from the decade following the year 2000. There is quite a bit of constructional variation involving this variant of “seeing the light” for realization⁶. The lightbulb can *go on* or *off in one’s head*, as in (7a-b), *over one’s head*, as in (7c-d), or be completely reduced to someone having *a lightbulb moment*, as in (7e-f).

- (7) a. *So, while I was flipping through these books, I suddenly had this little lightbulb go on in my head and I said, “Hey, I’ll write the kind of book I want my daughter to read”.* (SPOK: NPR_Weekend, 1994)
- b. *It’s funny that you bring up the apple because of all the topics in the book, I found the apple — every page was another lightbulb going off in my head, the aha factor in apples...that if you plant the seed from an apple that you’re eating, you’re not going to get that apple back again.* (SPOK: NPR_Science, 2001)
- c. *My reaction was — I believe a small lightbulb popped over the head, and I went, (singing) opportunity...* (SPOK: NPR_FreshAir, 2014)
- d. *I decided I was going to give her one last verbal command: “If you don’t drop that gun, I’m going to shoot you.” Finally, it’s like the lightbulb went on over this woman’s head. She dropped the gun.* (SPOK: NPR_FreshAir, 2004)
- e. *So here’s the big lightbulb moment for me. In 1994, someone got the idea of entering a group of Tarahumara runners in this legendary race called the Leadville Trail 100.* (SPOK: TEDRadioHour, 2015)
- f. *WINFREY: Thirty-two-year-old Glen says it took the loss of thousands of lives for him to have his own lightbulb moment. Take a look at what happened to Glen. Mr-GLEN-UPTON-1Pl: September 11th was a lightbulb moment for me.* (SPOK: Ind_Oprah, 2001)

⁶ A search in all of COCA for *PRO [see] the light* yielded 193 returns, 43 of which were figurative in the sense of ‘realize’. Of these, 17 involved a first person singular subject, 7 involved second person, 9 third person singular, and 10 third person plural. Again, we see that the expression is tilted towards a preference for 1sg subjects in terms of raw frequency. This matches the 1sg preference pattern for the figurative reading of *a/the lightbulb went off/on*, which collocates most frequently with *in my mind/head*, *over my head*, and the like.

Of the 24 figurative examples with *lightbulb* in COCA_{sp} indicating that ideation or realization is happening in or above someone’s head are no less than 11 separate constructions. Nevertheless, there is very interesting patterning by agreement, as shown in Table 10. One could even say that there’s a complementary distribution holding for the dative-like readings (in which a lightbulb goes on/off *for me*, but not for others), as well as expressions in which the lightbulb is explicitly located in *my* head, as opposed to being implicitly located in someone’s head).

Table 10

Observed frequencies for figurative uses of *lightbulb* in COCA_{sp}. We list the full range of ideational constructions-by-agreement patterns (where cognizer might be the subject (nominative), possessor (genitive) or beneficiary (dative) of the idea-emitting lightbulb. Although the frequencies are quite low, the inflectional skewing is quite pronounced

<i>lightbulb</i>	1SG	2	3SG	1PL	3PL	Row Total
<i>the lightbulb GO off (PRO implicit)</i>			2			2
<i>a lightbulb GO on (PRO implicit)</i>					2	2
<i>a lightbulb GO on over PRO’s head</i>			1			1
<i>a lightbulb POP over (PRO implicit)</i>	1					1
<i>the lightbulb GO off for PRO</i>	1					1
<i>the lightbulb GO off in PRO’s head</i>	2					2
<i>the lightbulb GO on in PRO’s head</i>	1					1
<i>PRO HAVE a lightbulb moment</i>		3	2			5
<i>PRO HAVE PRO’s lightbulb moment</i>		1	1			2
<i>BE a/the lightbulb moment for PRO</i>	5					5
<i>PRO’s lightbulb (moment)</i>	2					2
Column Total	12	4	6		2	24
Column Total / Row Total as Percentage	50%	17%	25%		8%	100%

Finally, we make mention of an idiomatic expression about sudden realization more prevalent in British English than in North American: *the penny dropped*. A search of the GLoWbE corpus (Corpus of Global Web-Based English) available through the BYU website (<https://corpus.byu.edu/glowbe/>) gives the following raw frequencies for the expression by country, where N ≥ 5: Great Britain (111), Ireland (32), Australia (32), USA (13), New Zealand (8), and Canada (5). There are only 2 examples in COCA_{sp}, but both make clear that the cognizer is 1SG, as shown in (8):

- (8) a. *I found the neurophysiology and the neuroanatomy the most interesting part of my studies, although it took a while before **the penny dropped** and I fell off my donkey and decided I was going to become a neurosurgeon*". (SPOK: NPR_Fresh Air, 2015)
- b. *Det-CHAMBERS: I didn’t know how David Coffin had died. No one knew. Ms-LEE: And that was the first time **the penny dropped**, and I went, “Oh, my God. Oh, my God”*". (SPOK: CBS_48Hours, 2007)

In the concluding section, we take stock of the semantic associations that tend to inhere in the verbal expressions about ideation surveyed here that disproportionately favour first person singular as opposed to second or third person cognizers.

5. DISCUSSION

Inspired by Sinclair 1991, Scheibman 2001, and other corpus linguists and grammaticalization scholars who advocate the importance of drilling down and examining the inflectional, agreement, and collocational preferences of individual verbs and verbal constructions, we have noted that 1sg and 2nd person ideation in English, prototypically associated with *I think* and *you know*, are each drawing in different kinds of expressions to do epistemic work in the language. Because the two prototypes are effectively functioning as pragmatic markers rather than complement-taking ideational verbs, at least in spoken varieties of the language, a host of other expressions from very different semantic fields are undergoing epistemization processes and entering into constructions about cognition. Not so surprisingly, those epistemizing expressions exhibiting a 1SG bias share many attributes not enjoyed, necessarily, by those expressions showing a bias towards 2nd person, much like THINK and KNOW have clearly gravitated in separate directions in terms of their agreement patterns. These differences lead us to conclude that 1SG and 2nd person (and possibly, 3rd as well) represent distinct styles of ideation and, consequently, have attracted and will continue to attract different kinds of novel expressions in their wake.

To put it in terms reminiscent of Spivey 2008 and contemporary cognitive scientists describing fluid models of categorization, high frequency of occurrence — be it in conceptualization pathways or motor routines — can be construed as “attractor basins” that act as centres of gravity for similar concepts or behaviours. Bybee (2010: 76—96) has similarly argued that forces of semantic change are affected by high-frequency items with heavy semantic pull. We, too, have applied this metaphor in an earlier presentation of this research, associating the Latin for ‘I think’, *cogito*, and the Latin for ‘you know’, *scis* (2SG) or *scitis* (2PL), with different cognitive models of ideation as if they were different craters on the moon. Admittedly, THINK and KNOW in the abstract both have an unbounded (in the sense of Langacker 1991: 85—91), atelic, and imperfective quality to them as event types describing cognition (compare them to REALIZE or LEARN). Nevertheless, we suggest that 1SG ideation, typified by *I think*, means something like ‘I (suddenly) realize something’, invoking semantic properties generally associated with prototypical transitive events, such as change of state, being telic and force-dynamic, and having a more compressed and punctual temporal profile. On the other hand, second person ideation, in the guise of *you know*, means something more stative like ‘you have a thought’ or ‘you (continuously) ponder/consider something’. Thus, its connotations could be characterized as less transitive, more atelic, more durative, and less likely to involve change over time. As high-frequency attractor basins representing the semantic field of cognition, *I think* and *you know* — or more succinctly, 1sg.ideate and 2.ideate — recruit different kinds of expressions to do epistemic work. Expressions examined above in COCA_{sp} like *what struck PRO about, it hit PRO that, a lightbulb went off in PRO’s head, it raced through PRO’s mind, the penny dropped (for PRO)*, show an undeniable preference for 1sg as PRO. These expressions overwhelmingly suggest a tight temporal profile and a discernable change of state or outcome; in short, a flash of realization. In a nutshell, when I think, my brain storms (*it struck me like a bolt of lightning, it came to me in a flash*); when you think, your brain waves (*something’s going on in your mind, you seem to be in the know*).

We would like to end with a caveat about the wider interpretation of our findings. Our study has limited itself to English data and, even then, the study has been largely limited to the usage of THINK and KNOW in the simple present tense and only in a conversational genre. Clearly, we are not in a position to make empirically justified claims about comparable behaviour of the counterparts to these verbs in other languages, or indeed other genres or tense/aspect categories in English. The SG/PL ambiguity associated with English *you* also invites further research into the preferences for 2SG and 2PL subject preferences for these verbs in other languages. There remains then the question of how specific to English our findings are and whether comparable preferences for subjects of THINK and KNOW occur sometimes, frequently, very frequently, or always in other languages. We believe these are questions that can and should be further explored.

© Sally Rice, John Newman, 2018

REFERENCES

- Agresti, A. (2007). *An Introduction to Categorical Data Analysis*, 2nd edition. Hoboken, NJ: John Wiley and Sons.
- Aijmer, K. (1997). *I think* — an English modal particle. In Swan, T & O. J. Westvik (eds.) *Modality in Germanic Languages: Historical and Comparative Perspectives*. Berlin/New York: Mouton de Gruyter, 1—47.
- Bybee, J. (2010). *Language, Usage, and Cognition*. Cambridge: Cambridge University Press.
- Dancygier, B. (2017a). Introduction. In Dancygier, B. (ed.) *The Cambridge Handbook of Cognitive Linguistics*. Cambridge: Cambridge University Press, 1—10.
- Dancygier, B. (ed.). (2017b). *The Cambridge Handbook of Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Davies, M. (2008). *The Corpus of Contemporary American English: 450 million words, 1990-present*. <http://corpus.byu.edu/coca/>.
- Goddard, C. (1997). *Semantic Analysis: A Practical Introduction*. Oxford: Oxford University Press.
- Goddard, C. & A. Wierzbicka. (2014). *Words and Meanings: Lexical Semantics Across Domains, Languages, and Cultures*. Oxford: Oxford University Press.
- Gries, S. Th. (2013a). *Statistics for Linguistics with R. A Practical Introduction*. Berlin: De Gruyter.
- Gries, S. Th. (2013b). Basic significance testing. In Podesva, R. J. & D. Sharma (eds.) *Research Methods in Linguistics*. Cambridge: Cambridge University Press, 316—336.
- Hale, K. L. (1972). A new perspective on American Indian linguistics. In Ortiz, A. (ed) *New Perspectives on the Pueblos*. Albuquerque: University of New Mexico Press, 87—103.
- Kärkkäinen, E. (2003). *Epistemic Stance in English Conversation: A Description of its Interactional Functions, with a Focus on I think*. Amsterdam/Philadelphia: John Benjamins.
- Knowles, G. & Z. M. Don. (2004). The notion of a “lemma”. *Headwords, roots and lexical sets*. *International Journal of Corpus Linguistics* 9, 69—81.
- Langacker, R. W. (1991). *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Berlin/New York: Mouton de Gruyter.
- Newman, J. & S. Rice. (2006). Transitivity schemas of English EAT and DRINK in the BNC. In Gries, S. Th. & A. Stefanowitsch (eds.) *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*. Berlin/New York: Mouton de Gruyter, 225—260.

- R Development Core Team. (2018). *R. A Language and Environment for Statistical Computing*. Vienna, Austria.
- Scheibman, J. (2001). Local patterns of subjectivity in person and verb type in American English conversation. In Bybee, J. & P. Hopper (eds.) *Frequency and the Emergence of Linguistic Structure*. Amsterdam & Philadelphia: John Benjamins, 61—89.
- Scheibman, J. (2002). *Point of View and Grammar: Structural Patterns of Subjectivity in American English Conversation*. Amsterdam/Philadelphia: John Benjamins.
- Schoonjans, S. (2012). The particularization of German complement-taking mental predicates. *Journal of Pragmatics* 44, 776—797.
- Silverstein, M. (1976). Hierarchy of features and ergativity. In Dixon, R. M. W. (ed.) *Grammatical Categories in Australian Languages*. Canberra: Australian Institute of Aboriginal Studies, 112—171.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Spivey, M. (2008). *The Continuity of Mind*. Oxford: Oxford University Press.
- Tao, H. (2001). Discovering the usual with corpora: The case of *remember*. In Simpson, R. & J. Swales (eds.) *Corpus Linguistics in North America: Selections from the 1999 Symposium*. Ann Arbor: University of Michigan Press, 116—114.
- Tao, H. (2003). A usage-based approach to argument structure: *Remember* and *forget* in spoken English. *International Journal of Corpus Linguistics* 8 (1), 75—85.
- Thompson, S. A. & A. Mulac. (1991). A quantitative perspective on the grammaticization of epistemic parentheticals in English. In Traugott, E. C. & B. Heine (eds.) *Approaches to Grammaticalization, Volume II: Focus on Types of Grammatical Markers*. Amsterdam/Philadelphia: John Benjamins, 313—329.
- Van Bogaert, J. (2011). *I think* and other complement-taking mental predicates: A case of and for constructional grammaticalization. *Linguistics* 49 (2), 295—332.
- Wierzbicka, A. (1996). *Semantics, Primes, and Universals*. Oxford: Oxford University Press.
- Wierzbicka, A. (2006). *English: Meaning and Culture*. Oxford: Oxford University Press.
- Wierzbicka, A. (2010). *Experience, Evidence, and Sense: The Hidden Cultural Legacy of English*. Oxford: Oxford University Press.

Article history:

Received: 10 May 2018

Revised: 25 May 2018

Accepted: 01 June 2018

История статьи:

Дата поступления в редакцию: 10 мая 2018

Дата принятия к печати: 01 июня 2018

For citation:

Rice, Sally and Newman, John (2018). A Corpus Investigation of English Cognition Verbs and their Effect on the Incipient Epistemization Physical Activity Verbs. *Russian Journal of Linguistics*, 22 (3), 560—580. doi: 10.22363/2312-9182-2018-22-3-560-580.

Для цитирования:

Rice, Sally and Newman, John. A Corpus Investigation of English Cognition Verbs and their Effect on the Incipient Epistemization Physical Activity Verbs // Вестник Российского университета дружбы народов. Серия: Лингвистика. = *Russian Journal of Linguistics*. 2018. Т. 22. № 3. С. 560—580. doi: 10.22363/2312-9182-2018-22-3-560-580.

Bionotes:

SALLY RICE is Professor in the Department of Linguistics at the University of Alberta. Her research spans multiple sub-disciplines, including lexical semantics, comparative Athapaskan, corpus linguistics, cognitive linguistics, and multimodality in language. She conducts fieldwork on Dene Sų́líné and Tsuut'ina, two northern Athapaskan languages, and has been an active proponent and practitioner of community-university research alliances — community-based corpus-building in particular. She was a co-founder of the Canadian Indigenous Languages and Literacy Development Institute (CILLDI), an annual summer school for speakers of indigenous languages, which trains community language activists in linguistic analysis, language pedagogy, and revitalization project development and advocacy. Contact information: sally.rice@ualberta.ca

JOHN NEWMAN is Professor Emeritus in the Department of Linguistics at the University of Alberta and an Adjunct Research Fellow in the School of Languages, Literatures, Cultures and Linguistics at Monash University. His research focus has been mainly in areas of cognitive linguistics, especially cognitive linguistic approaches to basic verbs. His research interests also include corpus linguistics, quantitative methods in linguistics, and historical linguistics. He has published research on Germanic, Sinitic, Austronesian, and Papuan languages. He is currently the Outgoing Editor-in-Chief of the journal “Cognitive Linguistics”. Contact information: john.newman@ualberta.ca

Сведения об авторах:

СЭЛЛИ РАЙС — профессор кафедры лингвистики в Университете Альберты. Сфера ее научных интересов охватывает многие направления, включая лексическую семантику, сравнительный анализ атабаскских языков, корпусную лингвистику, когнитивную лингвистику и мультимодальность в языке. Она проводит полевые исследования по двум северным атабаскским языкам — дене сулин и тсу т'ина — и является активным сторонником и организатором научного альянса сообщества общественности и университета, в частности в области совместного создания корпусов. Является одним из основателей Института канадских аборигенских языков и развития грамотности (CILLDI) и организатором ежегодной летней школы для говорящих на аборигенских языках, в которых проводятся обучение лингвистическому анализу, преподаванию языков, тренинги по развитию, защите и оживлению языков. Контактная информация: sally.rice@ualberta.ca

ДЖОН НЬЮМАН — Почетный профессор кафедры лингвистики в Университете Альберты и внештатный научный сотрудник Школы языков, литератур, культур и лингвистики в Университете Монаш. Его научные интересы находятся в области когнитивной лингвистики, в особенности подходы когнитивной лингвистики к базовым глаголам, они также включают историческую лингвистику, корпусную лингвистику, количественные методы исследований. Его публикации основаны на анализе германских, синитских, австронезийских и папуасских языков. Контактная информация: john.newman@ualberta.ca

ACKNOWLEDGMENTS

The authors contributed equally to the overall design of this research project and for the data collection and analysis. An earlier version of this research was presented at CSDL 12, UCSB, Santa Barbara, USA on November 4 2014, and we are grateful to the audience at that conference for valuable feedback. We would also like to thank Anna Gladkova and an anonymous reviewer for helpful comments on an earlier version of this article.

ФИНАНСИРОВАНИЕ И БЛАГОДАРНОСТИ

Авторы внесли равный вклад в написание данной статьи, а также в сбор и анализ материала. Предварительные результаты исследования были представлены на конференции CSDL12 (Conceptual Structure, Discourse, and Language), Санта-Барбара, США, 4.11.2014, и мы благодарны аудитории за участие в обсуждении. Мы также выражаем благодарность Анне Гладковой и анонимному рецензенту за прочтение нашей статьи и полезные комментарии.