1

A Few Words To Do With Multiword Expressions

PAUL KAY, LAURA A. MICHAELIS

Abstract

This paper provides a compositional, lexically based analysis of the infinitival, verb-headed idiom exemplified by the sentences *What does this have to do with me?* and *It may have had something to do with money*.[1] Using conventions of Sign-Based Construction Grammar (SBCG, Sag 2012, Kay and Sag 2012, Michaelis 2012), we show that this multiword expression is revealingly represented as an intransitive verb word *do*, whose subject cannot be locally instantiated, which is necessarily in base form, and which invokes or is invoked by other idiomatically construed lexemes, including a special subject-raising lexeme *have*, which contributes a (potentially null instantiated) degree argument. We argue that idiomatic *do*, despite its restricted combinatorial potential, is compositionally interpreted, denoting an association between two entities, the first of which is expressed by the non-locally-instantiated subject and the second of which is expressed by the *with*-headed PP. We draw several lessons from this study. First, as is perhaps self-evident, multi-word expressions that are composed mostly of idiomatic words, such as, *have, to,* and *do* in this idiom may also require the presence of non-idiomatic words. An example is the presence in the *to-do-*

[1] We are grateful for the privilege of contributing to a volume honoring Lauri Karttunen. Given that our contribution is about the form and meaning of an idiom, we find it a welcome coincidence that after a brilliant career that began in semantics and pragmatics, and then moved into computational linguistics, Lauri has recently returned to linguistic meaning, dissecting with his accustomed mastery a highly idiomatic class of raising adjectives with protean implicative properties, e.g., *You will be lucky to break even* (Karttunen 2013). We would also like to express our gratitude for the very helpful comments of an anonymous referee.

with idiom of ordinary, case-marking *with*. Second, there are idiomatic words that can occur in the expression but are not required. One such word is the idiomatic *have* of *have to do with*; there are nominal expressions containing *to do with* in the relevant meaning that lack *have*, such as *a question to do with money*. Finally, at a more general level, many multiword expressions are much like transparently interpreted word strings: they display the structures they do because of the combinatorial affordances of their lexical members; their phrasal properties are determined by the independently motivated and generally available phrasal constructions of the grammar.

1 Introduction

This chapter is centrally concerned with the idiom or multiword expression consisting of the words *to, do,* and *with*, usually in that order (1)[2], but not necessarily contiguous (2) and not always in that order (3).

(1) a. Inside the flat a fruity voice was reading out a list of figures which had something to do with the production of pig iron.
b. It may have had to do with money ...

(2) a. Anorexia is not just about losing weight it has to usually do with wanting control in your life.
b. As I see it, the decolonization movement in Indian country has to do mostly with countering "white" power with "Indian" power.
c. Walter said it had something complicated to do with quantum physics.
d. In both conceptions, boredom has to do fundamentally with an experience of time and problems of meaning.
e. What does that have to do with me?
f. What has that to do with me?

[2] All positive numbered examples in this chapter were attested on the web in January or February 2016, except as otherwise noted. In this chapter we use the terms 'idiom' and 'multiword expression' (MWE) interchangeably. We do not intend to imply thereby that we consider the terms indistinguishable in meaning, only that the linguistic expressions with which we are concerned here are aptly characterized by either term.

(3) a. Elinor needed little observation to perceive that her reserve was a mere calmness of manner with which sense had nothing to do.
b. We implemented a recurrent-descent parser, a scanner (with which I had little to do) and a runtime Basic environment.
c. ... and my limbs moved with a positiveness and precision with which I seemed to have nothing at all to do ...

The variation in linear order observed in (3) suggests one of the main points of the chapter – that those MWEs with fixed word order (e.g., *under the weather* (ill)*, in the pink* (healthy), etc.) are not representative: many, probably the majority, of MWEs are syntactically flexible, and so are best analyzed as bags of lexical items whose inherent, often idiosyncratically constrained, properties determine – in concert with the canonical lexical and phrasal constructions of the grammar – the meanings and syntactic distributions of the idioms (Nunberg, Sag & Wasow 1994, O'Grady 1998, Sag et al. 2002, Ramisch 2015, Karttunen 2013, Kay, Sag & Flickinger submitted). What holds this idiom, and many other idioms, together is not any sort of phrasal template, but the syntactic and semantic potentials of its constituent lexical items. We will see that the lexical pieces of this idiom display significant syntactic freedom while nonetheless obeying fairly strict and rather idiosyncratic syntactic limitations. Our argument will be that this particular syntax and the corresponding compositional semantics is accurately predicted by a perspicuous representation of the constituent lexical items against the background of the familiar, 'core', phrasal syntax of English. This approach is in the tradition of Nunberg, Sag, & Wasow (1994).

A second general point this idiom illustrates is that MWEs usually contain fewer words than are indicated for them in dictionaries and comparably authoritative sources. The only words that all tokens of the *to do with* idiom (hereafter TDW) contain are *to, do,* and *with*. Some versions of TDW listed in several online sources are given in (4). Some examples lacking either *have or be* and exhibiting non-quantifier determiners in the NP slot occupied by *something* in (4)a,b are given in (5). Moreover, as we will see below, the syntactic and semantic roles of *something* in (4)a,b and *The issue* in (5)a are distinct.

(4) a. be/have something to do with *sth* [Cambridge Advanced Learner's Dictionary]
b. be/have something to do with something [Macmillan Dictionary]
c. have to do with [Thesaurus.com]
d. have to do with [Collins English Dictionary]
e. have to do with [American Heritage Dictionary of the Idioms, Second Edition]

(5) a. The issue to do with liquidity is affecting this economy too much.
b. They asked a question to do with diversity.
c. I am attempting to recontact Brian about a question to do with my GE Profile microwave.
d. I had a question to do with the authenticity of painting...
e. It's an issue to do with dietary deficiency...

In (5)a, the TDW NP occurs as subject, in (5)b as object, and in (5)c object of a preposition. In none of the examples of (5) is the TDW phrase governed by *have* or *be*, nor does the nominal preceding *to do with* contain a degree-denoting expression such as *something, nothing,* or *a lot*. As we will see in the next section, TDW occurs in several different syntactic environments. We show below that Sign-Based Construction Grammar (SBCG, Sag 2012, Kay & Sag 2012, Michaelis 2012, Kay, Sag, & Flickinger submitted) yields a perspicuous analysis of the TDW idiom in terms of a small number of dedicated lexical items and, notably, no phrasal constructions.

Section 2 of this chapter lays out the essential data of TDW. Section 3 deals briefly with the version in which *be* replaces *have*. Section 4 provides a brief and general introduction to SBCG. Further detail of SBCG mechanics is introduced in the formal analysis of Section 6. An informal analysis of the TDW idiom is given in Section 5 and the SBCG version in Section 6. Section 7 concludes.

2 The Data

We are concerned with the TDW idiom having the sense of 'having connection with/being associated with in some way'. We are thus not concerned with examples like (6), on its least context-dependent interpretation.

(6) I'm looking for something to do with the kids today and you've given me some great tips!

There are three relevant cases, or sets of cases, to be considered.

Case 1. The most commonly noticed case of TDW consists of a VP headed by a verb *have* followed by a degree-quantifying or measure expression (*something, nothing, a great deal, totally, somewhat,* etc.) followed by a TDW infinitival verb phrase, as in (7).

(7) a. OK, so the kiss had something to do with the lace and the gauze and the tassels.
b. It has nothing to do with paying tithes. It has nothing to do with perfect Church attendance.
c. It may not have a great deal to do with trades union autonomy, but trades union autonomy may have a great deal to do with guiding the destinies of society, ...
d. This has nothing to do with Eric Burdon, but somewhat to do with Iceage.

Case 2. In the second, and also frequently noticed, case, TDW again appears as an infinitival VP complement to an intransitive verb *have* (or *be*, as discussed briefly in the next section), but not preceded by a measure phrase, as illustrated in (8).

(8) a. One of the biggest climate problems has to do with what most people eat.
b. This may have to do with factors of asymmetric neurologic development, such as being right or left-handed.
c. Slavery in ancient Rome and Greece didn't have to do with your race, like it did in America.

Case 3. The third case is less frequently recognized. In this case the TDW phrase is an infinitival relative clause, modifying the head of an NP that can subserve any of the normal functions of an NP. In (9) the NP containing the TDW infinitival relative clause functions, respectively, as subject in (9)a, object in (9)b, prepositional object in (9)c, and predicate in (9)d. In Case 3, the phrase immediately preceding *to* is necessarily nominal and it is not semantically restricted to degree quantification – see, e.g. (9)a,b – unlike the corresponding element in Case 1. Significantly there is no requirement that the TDW phrase be governed by *have* or *be* as well as no restriction to degree quantification.

(9) a. ... an issue to do with development of a railway station has been raging...
b. Well you've asked several questions to do with the refugee crisis.
c. Many times when I am outside I will be thinking about something to do with my work or a friend ...
d. ... he knew exactly what the issue was. It is a hardware problem and nothing to do with software. [3]

3 The *be* Version

For Cases 1 and 2 there exist variants in which a form of *be* occurs instead of the corresponding form of *have.*[4] .

(10) a. Cowardice is nothing to do with it - suicide takes considerable courage.
b. This new work is a lot to do with memory, childhood and innocence.
c. It could also be to do with the location of rooms...
d. The questions that were asked on 14 December last year were to do with looking into the assessment ...
e. What is that to do with us?

This usage appears to be a relatively recent, primarily British, innovation. It remains a minority usage in the British National Corpus and is quite rare in

[3] We note, but do not attempt to analyze further, a special use of idiomatic *do* that refers to a social association, as illustrated by (i).

(i) But they all rejected the proposal; and said, they would have nothing to do with me any more, neither on board, nor on shore...

In this use, TDW must be the complement of *have*, both the subject and complement of *with* must denote animate entities and the degree argument must be present (??*He was unwilling to have to do with me*). In addition, the social-association interpretation appears possible only in non-desiderative contexts, where idiomatic *have* is the complement of a verb like *want*: ??*He does not have anything to do with me*). Idiomatic *do* in this use appears to be a negative polarity item, requiring the presence of *not* or a negative degree argument: ??*He wants to have something to do with me*. Idiomatic *do* on its social association sense appears incompatible with extraction, although extraction contexts are generally viewed as non-veridical (negative polarity) contexts: ??*What would she have to do with you?*

[4] As noted, in Case 3, no particular verb governs the TDW NP.

contemporary American usage as registered in available corpora. The BNC yields 181 (34%) raw hits for combined *is/was/were/be to do with* and 352 (66%) raw hits for combined *has/had/have to do with*. The corresponding figures in the much larger Corpus of Contemporary American English are 44 (.6%) raw hits for combined *is/was/were/be to do with* and 7,350 (99.4%) raw hits for combined *has/had/have to do with*.[5] With respect to times of attestation, the OED lists examples of the *have* version from Middle English, while the earliest citation for the *be* version is 1902. We conclude that once we have specified a special lexeme *have* required to complete the analysis of Cases 1 and 2, to extend the analysis to the *be* usage requires no more than positing a corresponding *be* lexeme.

4 Sign-Based Construction Grammar

SBCG is based on a multiple-inheritance hierarchy of typed feature structures.[6] Readers acquainted with HPSG or with the kind of Berkeley Construction Grammar presented in, for example, Kay and Fillmore (1999), should find the tools and notations introduced in this paper familiar. This section does not go into details of notation. Those are explained, as necessary, as the SBCG analysis of TDW is presented in section 6.

The most important type of feature structure (FS) in SBCG is the *sign*, with subtypes *word*, *lexeme* and *phrase* (Sag 2012: 71). A *construct* is a local tree of signs. *Combinatorial constructions* define classes of constructs. *Lexical class constructions* can be thought of as filling in particular features of *listemes,* to license *lexemes*. A listeme is a minimal lexical entry that has not undergone derivational or inflectional elaboration (Sag 2012; 71, Di Sciullo & Williams 1987).

Each sign has a FORM feature, whose value is a morphological representation of the expression, notated here in standard English orthography. The other features of the sign are PHON(ology), ARG-ST (argument structure; for lexical signs only), SYN(tax), SEM(antics), and

[5] These figures are not corrected for meaning and doubtless exaggerate the advantage of *have* over *be* in both corpora because of the possession sense of *have* in, e.g., (6). We find no reason to suppose, however, that the strength of this artifact would be different in the two corpora.

[6] This section is based on Kay, Sag, & Flickinger (submitted). See Sag (2012) for a comprehensive introduction to SBCG.

CONTEXT.[7] The value of SYN is a feature structure that specifies the features CAT(egory), VAL(ence), and MRKG (marking). CAT values are FSs assigned to various word-class types (*noun, verb,* etc.), and specify values for the features appropriate to that type, including Lexical Identity (LID), whose value is a list of frames.[8] The MARKING (MRKG) value of a lexical sign (only) is normally identified with the LID value.[9] LID values are percolated from heads, making the identity of the lexical head available at every phrasal level; MRKG values are percolated from functors (i.e., specifiers or modifiers). Thus in an NP or N' the identity of both the head (via the LID value) and the modifier or specifier (via MRKG value) are visible to potential governing predicators.[10]

The value of the ARG-ST feature is a list of the *signs* that are the arguments – syntactic and semantic – of a predicating lexeme. Members of the ARG-ST list reappear in the list value of the VAL feature, except when extracted or given null realization. The ARG-ST feature is limited to lexical signs; unlike the VAL feature it does not appear in phrasal signs. This limitation constitutes one aspect of the highly local nature of SBCG. The external argument is the argument that gets 'raised', 'controlled', etc. When a lexeme has an external argument, it is the first element on the ARG-ST list. There is a syntactic CAT feature, XARG, whose value is the external argument.[11] The XARG feature in effect allows a predicator that selects a given type of phrasal sign to 'see' the first argument of the verb within that phrase. As well as figuring in control and raising, XARG feature is used to represent coreference constraints in more 'marked' constructions like copy raising, as discussed by Sag (2012: 142ff). A second property of the SBCG architecture that enforces locality is that only constructs, but not signs, have a DAUGHTERS (DTRS) FEATURE. Since the members of the list constituting the value of the ARG-ST feature are *signs* and since a *sign* does

[7] The PHONOLOGY feature is not relevant to the material of the present chapter and will not be included in our diagrams.

[8] Typically this will be a singleton list. See Sag (2012) for greater detail.

[9] Departures from the norm will be discussed. For readers familiar with HPSG, we note that the LID feature subsumes the function of the PFORM feature and accomplishes the analogous function for lexical heads of all categories.

[10] The MARKING feature is relevant to the analysis of idioms like *bark up the wrong tree*, in which not only must *up* be able to recognize that its complement NP is headed by *tree*, but critically that *tree* is modified by an AP headed by *wrong.* Since, analysis of the TDW idiom does not involve the MARKING feature, that feature is not systematically included in the diagrams of this chapter.

[11] For convenience, we sometimes refer, somewhat loosely, to the value of the XARG feature as 'the XARG'.

not have a DTRS feature, it is possible for a predicator to specify properties only of its daughters, not of its 'granddaughters'. Long-distance dependencies are encoded as a sequence of local dependencies, via the GAP (aka SLASH) feature, in the tradition of GPSG and HPSG.

Like the ARG-ST feature, the VAL feature takes as its value a list of signs, corresponding to the elements that an expression can combine with: the syntactico-semantic arguments of the predicator in order of increasing obliqueness, or decreasing accessibility (subject, direct object,...). Expressions including NPs, PPs, APs, and clauses have the empty list as their VAL value: they are 'saturated', in the sense that they already contain all of the predicator's arguments. As mentioned, the elements of the VAL list of a lexeme or word are the same signs as those appearing on the ARG-ST list, minus those that are extracted (encoded on the GAP list) or subject to null instantiation, definite or indefinite.

The value of SEM includes specifications for the features INDEX and FRAMES. We assume an indefinitely large number of referential indices *1, 2, ...* The FRAMES feature takes a list of elementary predications[12] as its value. We adopt a frame-based conception of semantics (Fillmore 1982, 1985, Fillmore and Baker 2010). We assume further that frames form a multiple-inheritance hierarchy in which there is a first-cut bifurcation of the universe of frames into canonical frames (*c-frame*s) and idiomatic-frames (*i-frames*). We assume, for example, that the idiomatic *do* of TDW means something like 'have some relation with' or 'be connected somehow with', and that its PP[*with*] complement is of the case-marking (semantically non-contributing) variety. Hence the listeme for the *do* of TDW is assigned [SEM [FRAMES *<i(diomatic)-do[rel(ation)]-fr(ame)>*, the material in *[brackets]* expressing a rough gloss or other indication of the meaning or function of the idiomatic word. Except in the case of semantically empty lexemes like the *to* and *with* of TDW, the LID value of a lexeme is identified with its FRAMES value.

Semantically null expressions, such as the idiomatic *to* of TDW, receive the specification [SEM [INDEX *none*, FRAMES < >]]. In the case of such semantically empty lexemes, the LID value is a *non-empty* list containing a

[12] In the present study these lists are all singleton (See Sag 2012). Roughly, SBCG frames are equivalent to the RELS of Minimal Recursion Semantics (MRS), in Copestake et al. (2005). We adopt here a simplified form of MRS, making no use of either the LTOP feature or the mechanisms for keeping track of relative scoping relations. We make limited use of the LABEL feature, which is explained where first employed.

semantically null frame, which serves to identify the lexical head when it is projected to higher levels of the phrase. Accordingly, despite the fact that *to* has an empty FRAMES value, we assign it a non-empty LID value: [SYN [CAT [LID <*i-to[TDW]-fr*>]]. In this manner, the *have* of TDW can identify a VP headed by TDW *to;* the *to* of TDW will be able to identify a VP headed by TDW *do* because of the latter's specification [SYN [CAT[LID <*i-do[rel]-fr*>]]]; and TDW *do* will be able to subcategorize for a PP headed by case-marking *with* because the latter will carry the specification SYN [CAT [LID <*with[null]-frame*>]][13].

The basic mechanism for preventing idiom words – like *beans* in *spill the beans* or *do* in TDW – from appearing in contexts where they are not lexically governed by the appropriate idiom predicator involves the feature VAL. Canonical (non-idiomatic) predicators are lexically specified as requiring all members of their VAL list to be non-idiomatic, that is, [LID <*c-frame*>]. By contrast, an idiomatic predicator such as idiomatic *spill* requires a direct object that is idiomatic, i.e. one whose LID value (and hence that of its lexical head) contains a particular *i-frame*, to wit: *i-beans[secret]-fr*. Similarly, the VAL value of the infinitive-taking verb *to* of TDW specifies a VP complement whose LID value is <*i-do[rel]-frame*>.

We will have occasion below to talk about the verb *have* that plays a role in the TDW idiom. We will discover that this verb, like the TDW infinitive *to,* is a subject-raising verb that contributes nothing to the semantics. Although this *have* is idiomatic in the sense that it plays a role in a MWE, the LID of an idiom-governing verb like TDW *have* does not contain an *i-frame* because a phrase headed by such a predicator can be a complement of an ordinary predicator. This circumstance is exemplified by the *spill* of the *spill the beans* in (11)a and by the *have* of TDW in (11)b and (11)c.[14]

(11) a. We both began to spill the beans about our relationships...

b. Some historians believe it has something to do with handicrafts that the first Strigolniki were engaged in ...

---

[13] We treat infinitive *to* as a highly defective auxiliary verb in the phrase structure grammar tradition (Gazdar, Pullum & Sag 1982, see also Pullum 1981).

[14] The *to* of TDW does have an *i-frame* LID because it must be recognizable to TDW *have.* It is not the case, however, that any infinitive-marking verb *to* that can take a VP idiom complement has a special form restricted to that complement.

c. Others believe it has to do with how you got your financing or who distributed your film.

In (11)b and (11)c the complement of *believe* is a clause headed by TDW *have,* illustrating the fact that the highest predicator in an MWE has to bear a *c-frame* LID.

5 The Analysis

In this section we provide an informal description of our analysis of TDW. In the next we show how this analysis is expressed more formally in SBCG.

5.1 Case 1

We repeat in (12) an example of Case 1 of the TD idiom, given above as (7)a.

(12) OK, so the kiss had something to do with the lace and the gauze and the tassels.

As mentioned, in Case 1, when an element intervenes between TDW *have* and TDW *to*, it must denote a degree-measuring expression, such as *something, a little, little, plenty, a great deal, somewhat,* etc. In a sentence like (13)a, in which the NP intervening between *have* and *to* is not a degree-quantifier, *it*-cleft (version b) and *wh*-cleft (version c) are possible. However, in a sentence genuinely illustrating Case 1, in which the intervening element is a degree-quantifying expression, neither *it*-cleft nor *wh*-cleft are possible, for reasons we discuss below. These facts establish that the *have* of (13)a,b,c are the ordinary 'possessive' have, not the special *have* of TDW. Example (13)d exemplifies Case 1, and the (13)e (*it*-cleft) and (13)f (*wh*-cleft) examples show that in Case 1 the string corresponding to *something to do with your computer* is not a constituent.

(13) a. if any of you guys has a problem to do with his computers [please keep in touch and we can solve [i]t].
b. If it's a problem to do with his computer that any of you guys have... (invented)
c. If what any of you guys have is a problem to do with his computer... (invented)
d. If your problem has something to do with your computer we can solve it. (invented)
e *If it's something to do with your computer that your problem has we can solve it. (invented)
f *If what your problem has is something to do with your computer we can solve it. (invented)

It follows from the fact that Case 1 examples require the intervening element to be a degree-quantifier that cases in which a non-degree-quantifier intervenes, like (13)a-c, are instances of the familiar 'possessive' have that occurs in a sentence like (14)a or (14)b.

(14) a. I have promised Lord Henry to go with him.
b. She was much affected, and promised immediately to set about the great work of seeking the salvation of her soul.

Similarly, the object of possessive *have* can be topicalized (15)a, but not the superficially similar [*something to do with* NP] string in a Case 1 sentence like (15)b, as shown in (15)c.

(15) a. A problem with a computer Einstein would never have. (invented)
b. Einstein's problem would never have something/anything to do with a computer. (invented)
c. *Something/Anything to do with a computer Einstein's problem would never have.

In our proposed analyses of Case 1 sentences, *have* is a subject raising verb with a non-subject argument directly following the verb that fails to have all the properties of a direct object. In this regard *have* functions like *promised* in (14).

Turning now to the semantics of Case 1, in a sentence like (16), the modern Nahuatl language is asserted to have an association with the language of the Aztecs and 'only a little' is asserted to express the strength of that association. The latter effect is achieved, as we will see below, by 'only a

little' binding a degree variable in the *i-do[rel]-frame* that idiomatic *do* introduces into the semantics of the sentence.

(16) Today, the modern Nahuatl language has only a little to do with the language spoken by the Aztecs.

5 Case 2

The relevant examples, like those in (8), the first of which is repeated as (17), contain the verb *have* and lack a degree quantifier between *have* and *to*.

(17) One of the biggest climate problems has to do with what most people eat.

These examples show that the degree-quantifier of Case 1 is an optional argument of *have* syntactically and that semantically the degree of association can simply remain unspecified. This observation comports with the fact that pairs of sentences like (18)a and (18)b appear to be indistinguishable in denotation.

(18) a. It had something to do with love.
b. It had to do with love.

As mentioned TDW *to* subcategorizes for a VP headed by TDW *do*, which in turn subcategorizes for ordinary case-marking PP[*with*]. Since *to* is a semantically inert subject-raising verb, like TDW *have*, in (17)a the denotatum of *One of the biggest climate problems* is identified with an entity that is associated (to an unspecified degree) with the denotatum of *what most people eat*.

5.3 Case 3

In this case the word-string [NP_1 *to do with* NP_2] is analyzed as a freestanding NP, headed by NP_1, which can subserve any of the normal functions of a referential or predicative NP, as illustrated in (9); (9)a is repeated as (19).

(19) ... an issue to do with development of a railway station has been raging ...

The string [*to do with* NP] is analyzed as an infinitival relative clause, modifying, for example, *issue* in (19).[15]

6 The Analysis Implemented in SBCG

The challenging problem for the analysis of TDW is to formulate the listemes for *do* (and to a lesser extent *to,* and *have*) in such a way that the TDW phrase can serve as a subject-raising verbal complement (Cases 1 and 2) and an infinitival relative clause (Case 3), and yet not license further imagined possible syntactic configurations, such as a finite clause like (20) a, a modifying gerund phrase like (20)b, a *for...to* infinitival clause, as in (20)c, a TDW infinitival VP governed by a subject raising verb other than *have,* as in (20)d, and so on.

(20) a. *The problem does with fleas. 'The problem has to do with fleas'
b. *I have a problem doing with my kitchen. 'I have a problem to do with my kitchen'
c. *Her mother doesn't want for her to do with him anymore. 'Her mother doesn't want her to have to do with him anymore.'
d. *This film appears to do with the war. 'This film has to do with the war.'

[15] It is tempting to assume that infinitival relative clauses attach to NP. Sag 1997 argues convincingly that *wh*-relatives attach to NP. He argues further, though, that non-finite relatives attach to N', in order to account for the fact that bare relatives precede *wh*-relatives. However, simple infinitival relatives do not necessarily precede *wh*-relatives, as seen in (i).

(i) I need someone who I can trust to talk to.

We note that a saturated nominal word, such as *something,* might qualify as an NP in the intended sense of having an empty VALENCE list (See also Sag 2012: 80, fn. 22), although Sag (1997: 36-37 note) suggests analyzing words like *something* as *some thing*, which would support the proposal that infinitival relatives attach to N'. We will not attempt to settle that matter here, nor will we present an SBCG formulation of the construction(s) licensing infinitival relative clauses. For detailed analysis of English relative clauses in an early form of constructional HPSG, see Sag 1997.

We will argue that SBCG, as represented in Boas & Sag (2012) and Kay, Sag, & Flickinger (submitted) makes possible a perspicuous implementation of the analysis sketched in this section.

6.1 *with*

Case-marking *with* is not an idiomatic lexeme restricted to the TDW idiom. However, it provides a good starting point for our analysis, both because it is required in the TDW idiom and because case-marking *with* furnishes a convenient opportunity to introduce some of the notation of SBCG. The lexical entry ('listeme') for *cm-with* is given in (21).

(21)
$$\left[\begin{array}{ll} \textit{cm-p-word} & \\ \text{FORM} & \langle \textit{with} \rangle \\ \text{ARG-ST} & \langle \textit{Y}\text{:NP}_{y} \rangle \\ \text{SYN} & \left[\begin{array}{ll} \text{CAT} & \left[\begin{array}{ll} \textit{prep} & \\ \text{XARG} & \textit{none} \\ \text{LID} & \textit{with[cm]-fr} \end{array}\right] \\ \text{VAL} & \langle \textit{Y} \rangle \end{array}\right] \\ \text{SEM} & \left[\begin{array}{ll} \text{INDEX} & \textit{y} \\ \text{FRAMES} & \langle\ \rangle \end{array}\right] \end{array}\right]$$

The notation in italics at the top of the AVM indicates a case-marking prepositional word. The second row tells us the morphological form of this word. Lists are surrounded by <angled brackets>. ARG-ST abbreviates *argument structure*. We see that this word has a single argument, an NP whose index is *y*, and which is tagged with the feature-structure tag italic capital *Y*. The SYN value is represented by an AVM containing the features CAT and VAL. The CAT value is a FS of type *preposition* with external argument (XARG) indicated as *none* – consonant with the 'case-marking' function – and whose lexical identity (LID) identifies this word, and any phrase it projects, as headed by case-marking *with*. The VAL value is a singleton list containing the tag *Y*, identifying it with the single member of the ARG-ST list. The SEM value shows that the referential INDEX of this word, which will be projected to the INDEX of the PP it heads, is identified with that of its unique (object) argument. This fact, along with the fact that

the FRAMES list is empty, completes the specification of a preposition with 'case-marking' function.[16]

We now want to consider the structure of a PP projected by case-marking *with*. The combinatorial construction that assembles a case-marking preposition, one which makes no semantic contribution and has no XARG, with its NP object is the Saturational Head-Complement Construction (Sag 2012: 152), shown in (22) The construction is called 'saturational' because there is no XARG that needs to be controlled or realized outside of the PP.

(22) **Saturational Head-Complement Construction** (↑ *headed-construct*)

$$\textit{sat-hd-comp-cxt} \Rightarrow \left[\begin{array}{ll} \text{MTR} & \left[\text{SYN } X!\left[\text{VAL } \langle\ \rangle\right]\right] \\ \text{DTRS} & \langle Z \rangle \oplus L\text{:}\textit{nelist} \\ \text{HD-DTR} & Z\text{:}\left[\begin{array}{ll} \textit{word} & \\ \text{SYN} & X\text{:}\left[\begin{array}{ll} \text{CAT} & \left[\begin{array}{ll} \textit{prep} & \\ \text{XARG} & \textit{none} \end{array}\right] \\ \text{VAL} & L \end{array}\right] \end{array}\right] \end{array}\right]$$

The statement in (22) defines the type *saturational-head-complement-construct*, which is the class of FSs named by the type designation preceding the double-shafted arrow and constrained by the feature-structure description expressed in the AVM following the arrow. The notation (↑ *headed-construct*) indicates that the type *saturational-head-complement-construct* is a subtype of *headed-construct* (SAG 2012: 154). A *construct* is a FS that defines a local tree, a mother immediately dominating one or more daughters. In the AVM there is a mother (MTR) feature, a daughters (DTRS) feature, and a head daughter (HD-DTR) feature. By definition, the mother and each of the daughters are of type *sign*, and since the *sat-hd-comp-cxt* type is a subtype of *headed-construct,* instances of *sat-hd-comp-cxt* inherit the constraints defined (in analogous fashion) for *headed-construct*[17]. In the mother's SYN value the first thing encountered is the compound tag symbol '*X*!'. In the head daughter's SYN value we find the compound tag symbol '*X*:'. Pairs of elements so tagged (one ending in '!' and one either ending in ':' or lacking a punctuation mark) are identical except for the parts in which they are explicitly shown to differ (Sag 2012:126 note). In this case, the SYN values of the MTR and HD-DTR are

[16] Situations in which the FRAMES value is the empty list constitute the only circumstance considered in this chapter in which the LID and FRAMES values are not identified, Cf. fn. 9.

[17] We do not show the details here. For more on the headed constructions and on the formal nature of the SBCG type hierarchy, see Sag (2012).

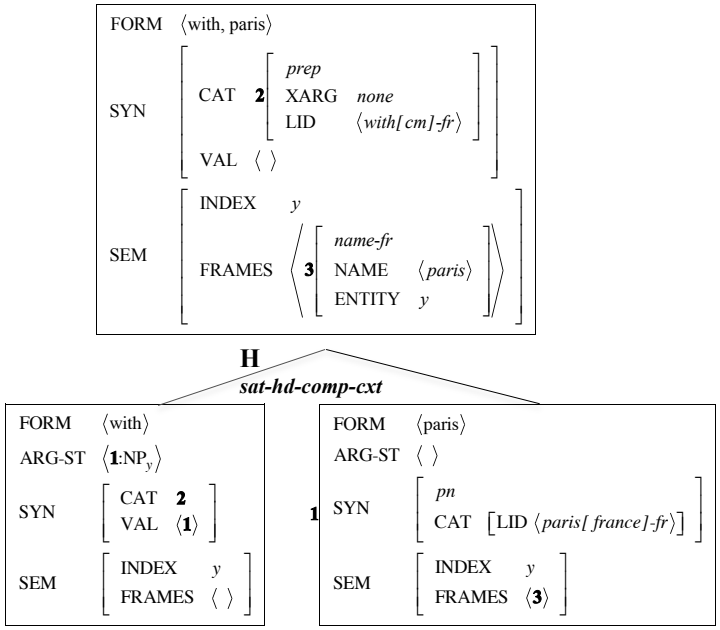
identical, except that the MTR's VAL list is empty. The HD-DTR's CAT feature indicates a preposition with no XARG, a case-marking preposition. The DTRS feature is a list consisting of the head daughter (note the *Z* tag) followed (as indicated by the symbol '⊕') by a non-empty list, which is identified (via the list tag *L*) with the VAL list of the head daughter.

The listeme for *Paris,* following Sag's (2012: 108-109) treatment of proper names, is given in (23). The construct *with Paris* – licensed by the listeme for case-marking *with* (21), the Saturated Head-Complement Construction (22), and the listeme *Paris* – is given in (24). [18]

(23)
$$\left[\begin{array}{ll} \textit{word} & \\ \text{FORM} & \langle \text{paris} \rangle \\ \text{ARG-ST} & \langle\ \rangle \\ \text{SYN} & \left[\begin{array}{l} \textit{pn} \\ \text{CAT}\ \left[\text{LID}\ \langle \textit{paris [france]-fr} \rangle\right] \end{array}\right] \\ \text{SEM} & \left[\begin{array}{ll} \text{INDEX} & y \\ \text{FRAMES} & \left\langle \boxed{3} \left[\begin{array}{ll} \textit{name-fr} & \\ \text{NAME} & \langle \textit{paris} \rangle \\ \text{ENTITY} & y \end{array}\right] \right\rangle \end{array}\right] \end{array}\right]$$

[18] FSs are enclosed in $\boxed{\textit{boxes}}$. FS descriptions are enclosed in [square brackets].

(24)



The bold **H** on the left descending branch indicates the head daughter ([FORM <with>]), whose single argument (**1:**NP_y) is also its unique valent (**1).** As a case-marking preposition *with*'s SEM value shows an empty FRAMES list and its INDEX (*y*) is identified with that of its object. The non-head daughter ([FORM <paris>]) has the internal structure specified in (23).[19] The mother constituent ([FORM <with,paris>]) inherits its SYN|CAT value (**2**) from the head daughter. According to the Saturated Head Complement Construction, its VAL value is the empty list, since the unique valent of the head daughter is realized as the non-head daughter (**1**). The mother's INDEX (*y*) is inherited from the head daughter. Since a mother constituent inherits the frames of all its daughters according to the Principle of Compositionality (Sag et al. 2003) and since the FRAMES list of *with* is empty, the only frame in the mother's LIST is the unique frame (**3)** of the non-head daughter**.**

Starting with the phrase *with Paris*, shown as the top box in (24), proceeding word by word we build up the VP *do with Paris*; then the VP *to do with Paris*; then the VP *have to do with Paris*, exemplifying Case 2; and

[19] Since we are dealing here with a specific token, modeled by a FS, rather than a class of FS as in (23), the AVM is enclosed in a box instead of square brackets

the VP *has something to do with Paris,* exemplifying Case 1. Subsequently, we address the derivation of the NP *something to do with Paris*, exemplifying Case 3, in which *to do with Paris* is not a VP but an infinitival relative clause..

6.2 *do*

The special word *do* is at the heart of the TDW idiom. The listeme for TDW *do* is shown in (25).

(25)

intrans-v-word		
FORM	⟨do⟩	
ARG-ST	⟨*X*:NP$_x$, *Y*:PP$_y$[LID ⟨*with[cm]-fr*⟩]⟩	
SYN	CAT	LID ⟨Z⟩
		XARG *X*
		VF *base*
	VAL	⟨*X*,*Y*⟩
SEM	FRAMES	⟨ Z: [*i-do[rel]-fr* ; DEGREE *index* ; TRAJECTOR *x* ; LANDMARK *y*] ⟩

TDW *do* is an intransitive verb word with the morphological form *do*, as shown in the top two rows of the AVM. This intransitive verb word finds a precedent in intransitive uses of *do* like *This will do* and *She is doing just fine*. Its ARG-ST list has two members, an NP with INDEX *x* and a PP with INDEX *y*, identified as the case-marking PP[*with*] by its LID value *<with[cm]-fr>*.[20] Skipping for a moment to the semantics in the bottom row, we find a single frame, *i-do[rel]-fr,* expressing the information that a TRAJECTOR *x* bears an unspecified scalar relation to a LANDMARK *y* of a DEGREE to be determined. Thus the values of the TRAJECTOR and LANDMARK semantic arguments are identified with the respective indices of the NP and PP syntactic arguments. The CAT value shows this verb word to be of *base* form, which eliminates (20)a,b, repeated as (26).

[20] Paths in AVM's may be abbreviated. In (25), [SYN [CAT [LID <*with[null]-fr*>]]] is rendered [LID <*with[null]-fr*>].

(26) a. *The problem does with fleas. 'The problem has to do with fleas'
b. *I have a problem doing with my kitchen. 'I have a problem to do with my kitchen'

The XARG value is identified with the NP_x argument and the LID value is identified with the FRAMES value, as is canonical. Examples like those in (27) are ruled out by the fact that TDW *to* will be assigned the *i-frame i-to[TDW]-fr* and the only predicators that can accept as a complement a constituent with [LID <*i-to[TDW]-fr*>] are TDW *have* and *be*.

(27) a. *Her mother doesn't want for her to do with him anymore. 'Her mother doesn't want her to have to do with him anymore.' (= (20)c)
b. *I expect the film to do with the war. 'I expect the film to have to do with the war.'

The TDW VP *do with Paris* is licensed by the word *do*, the PP *with Paris*, and the Predicational Head-Complement Construction (28), which licenses VPs, along with other head-complement phrases such as canonical PPs, and certain APs, and N-bars. The Predicational Head-Complement Construction differs from the Saturational Head-Complement Construction in that it projects from heads with non-empty VAL values. The Predicational Head-Complement Construction appears in (28).

(28) **Predicational Head-Complement Construction** (↑ *headed-construct*)

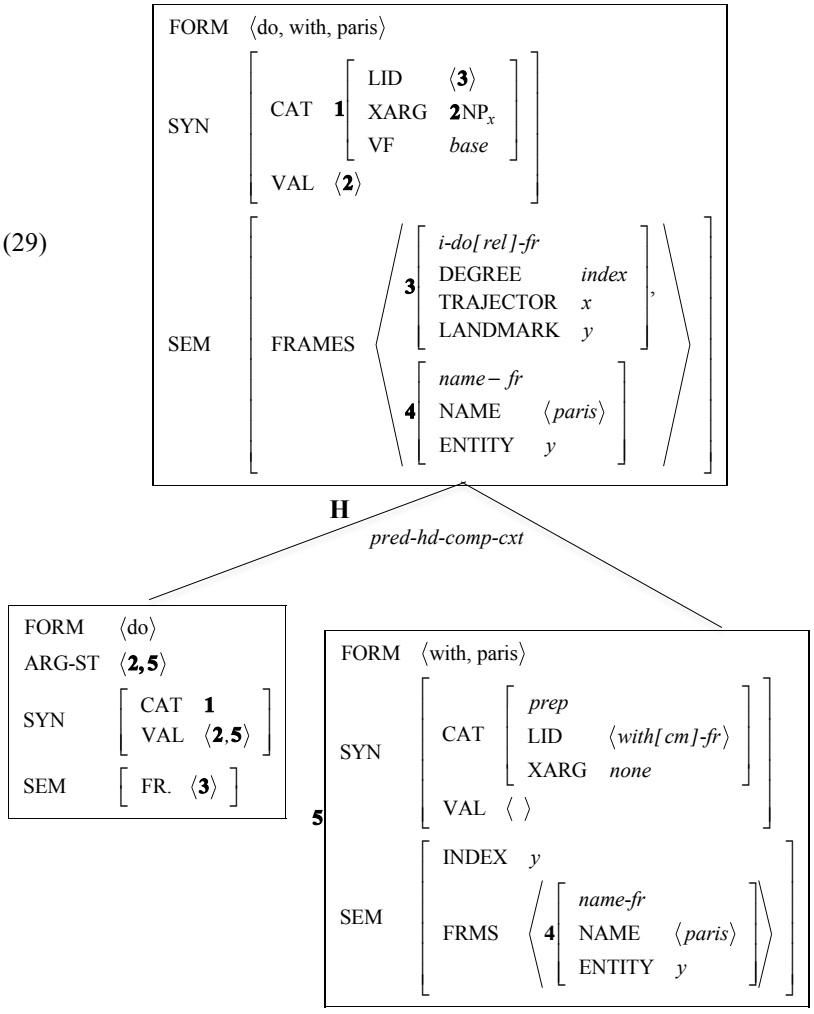
$$\textit{pred-hd-comp-cxt} \Rightarrow \left[\begin{array}{ll} \text{MTR} & \left[\text{SYN } X\text{:}\left[\text{VAL } \langle Y \rangle\right]\right] \\ \text{DTRS} & \langle Z \rangle \oplus \textit{L:nelist} \\ \text{HD-DTR} & Z\text{:}\left[\begin{array}{ll} \textit{word} & \\ \text{SYN} & X\text{:}\left[\begin{array}{ll} \text{CAT} & [\text{XARG } Y] \\ \text{VAL} & \langle Y \rangle \oplus L \end{array}\right] \end{array}\right] \end{array}\right]$$

The construction in (28) defines the type *predicational-head-complement-construct* as the class of constructs that obey the constraints depicted in the AVM to the right of the arrow. There is a mother feature MTR, a daughters feature DTRS, and a head daughter feature HD-DTR. The mother's and head daughter's syntax differ only in the facts (1) that the mother, like all phrases, lacks an ARG-ST feature and (2) that the mother's VAL list consists only of the XARG (*Y*). The other members of the daughter's VAL

list, indicated by the list tag *L*, are realized as sisters to the head daughter, as displayed in the value of the DTRS feature. The DTRS list consists of the head daughter followed by its non-XARG complements.

The construct whose mother is the TDW VP *do with Paris* appears in (29).

(29)



In construct (29), the mother, *do with Paris,* inherits its CAT value (**1**), from the head daughter *do,* specifying [LID <*i-do[rel]-fr*>] (**3**)**,** [XARG NP$_x$] (**2**), and, [VF *base*]. The two members of the mother's FRAMES list, *i-*

do[rel]-fr (**3**) and *name-fr* (**4**) are inherited from the head and non-head daughters, respectively.[21] Regarding the arguments of the head daughter, the tag **2** identifies the XARG of the mother, and the tag **5** identifies the unique valent of the head daughter – consequently, also the non-head daughter.

6.3 *to*

TDW *to* has both a verb and a complementizer version, the former heading the VP in cases 1 and 2 and the latter heading the infinitival relative clause in case 3.[22] We consider the verb TDW *to* in this section and complementizer TDW *to* later. The TDW verb *to* shaes many of the properties of the common infinitival verb *to*. It is nevertheless unique in that (1) TDW *to* requires specifically that its base-form VP complement be specified [LID <*i-do[rel]-fr*>], and (2), TDW *to* bears an idiomatic LID value, <*i-to[TDW]-fr*>, which ensures that a VP headed by TDW *to* can only serve as complement for a verb that specifies this LID in its VAL list, the unique verbs with this property being TDW *have* and *be*. In (30) we show the SELECT value as *none*, which contrasts with the complementizer version, whose SELECT value will be non-null. The listeme for the verb TDW *to* appears in (30).

(30)
$$\left[\begin{array}{ll} \textit{sraising-verb-word} & \\ \text{FORM} & \langle \textit{to} \rangle \\ \text{ARG-ST} & \left\langle \begin{array}{l} X\text{:NP}, \\ Y\text{:VP}\left[\text{LID } \langle \textit{i-do[rel]-fr} \rangle\right] \end{array} \right\rangle \\ \text{SYN} & \left[\begin{array}{ll} \text{CAT} & \left[\begin{array}{ll} \text{LID} & \langle \textit{i-to[TDW]-fr} \rangle \\ \text{XARG} & X \\ \text{SELECT} & \textit{none} \\ \text{VF} & \textit{inf} \end{array}\right] \\ \text{VAL} & \langle Y \rangle \end{array}\right] \\ \text{SEM} & \left[\begin{array}{ll} \text{INDEX} & \textit{none} \\ \text{FRAMES} & \langle\ \rangle \end{array}\right] \end{array}\right]$$

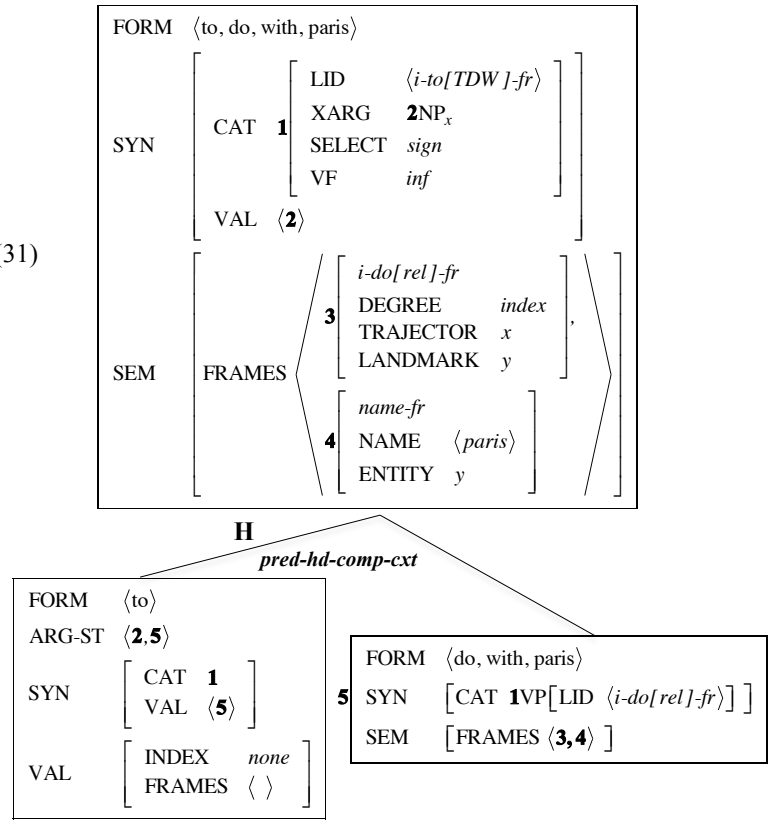
Since TDW *to* is shown at the top of the AVM to be of the type *s(ubject)raising-verb-word* no further indication of the identity of the

[22] Verbs and complementizers share many properties and are grouped together in the type hierarchy of SBCG as the two immediate subtypes of the type *verbal* (SAG 2012: 81-82).

indices of the XARG and the complement VP's XARG valent are displayed in diagram (30). The ARG-ST shows an NP XARG and a VP complement with [LID <*i-do[rel]-fr*>]. The CAT value shows the VERB FORM (VF) to be *inf,* the XARG to be identified with the NP argument, and the LID to contain the idiomatic frame *i-to[TDW]-fr*, which ensures that a VP headed by TDW *to* can complement only TDW *have* (or *be*)..

The VP *to do with Paris*, is licensed by TWD *to* (30), the Predicational Head-Complement Construction (28), and the VP *do with Paris* the mother of (29). The construct whose mother is the TDW VP *to do with Paris* is given in (31).

(31)



In the mother FS of (31), the phrase *to do with Paris*, the CAT value (**1**) is identified with that of the head daughter *to* – exemplifying a property of all

headed phrases – and includes the specifications [VF *inf*], [XARG NP_x] and [LID <*i-to[TDW]-fr*>]. The mother's frames, *i-do[rel]-fr* (**3**) and *name-fr* (**4**), are inherited from the non-head daughter *do with paris* (**5**), which satisfies the unique VAL requirement of the head daughter, *to*.

6.4 *have*

The listeme for TDW *have* is shown in (32).

(32)
$$\left[\begin{array}{ll}
\textit{sraising-v-lxm} & \\
\text{FORM} & \langle \text{have} \rangle \\
\text{ARG-ST} & \left\langle \begin{array}{l} X\text{:NP}, \\ Y\text{:XP}\left[\text{FRAMES}\left\langle \ldots\left[\begin{array}{ll}\textit{deg-quant-fr} & \\ \text{MEASURE} & w\end{array}\right]\ldots\right\rangle\right], \\ Z\text{:VP}\left[\left[\text{LID}\ \langle \textit{i-to[TDW]-fr}\rangle\right]\left[\text{FRAMES}\left\langle \ldots\left[\begin{array}{ll}\textit{i-do[rel]-fr} & \\ \text{DEGREE} & w \\ \ldots & \end{array}\right]\ldots\right\rangle\right]\right] \end{array} \right\rangle \\
\text{SYN} & \left[\begin{array}{ll}\text{CAT} & \left[\begin{array}{ll}\textit{verb} & \\ \text{LID} & \langle \textit{have[TDW]-fr}\rangle \\ \text{XARG} & X\end{array}\right] \\ \text{VAL} & \left\langle X, Y^{ini}, Z\right\rangle\end{array}\right] \\
\text{SEM} & \left[\begin{array}{ll}\text{INDEX} & \textit{none} \\ \text{FRAMES} & \langle\ \rangle\end{array}\right]
\end{array}\right]$$

TDW *have* licenses examples like those in (18) repeated as (33)a, exemplifying Case1, and (33)b, exemplifying Case 2.

(33) a. It had something to do with love
b. It had to do with love.

Like TDW *to*, TDW *have* is a subject-raising verb. Its ARG-ST specifies, in addition to an XARG NP (*X*) and an infinitival VP complement (*Z*), an additional measure-phrase argument (*Y*). The *Y* argument contains a degree-quantifying frame whose MEASURE value *w* binds the value of the DEGREE argument of the *i-do[rel]-frame* of the *Z* argument. In the CAT value, the XARG status of the *X* argument is expressed and the LID value is

shown to be a canonical frame, since a phrase head by TDW *have* can be embedded under a canonical predicate, as in (34).

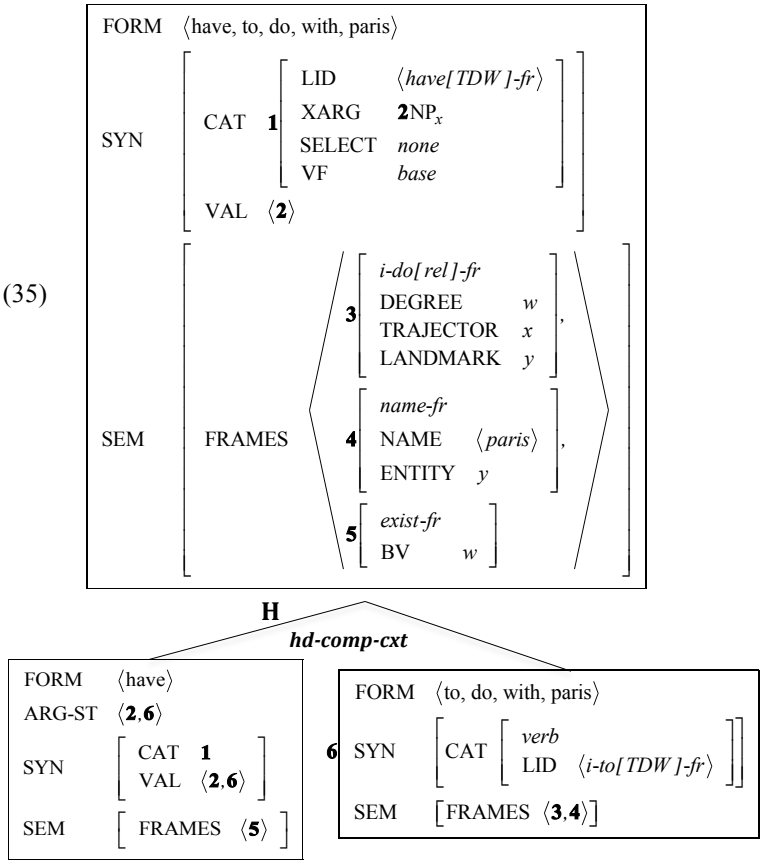
(34) His rationale was that this area of the spine innervated the pancreas, which was believed to have something to do with diabetes.

All three arguments appear in the VAL list. However, the second, *Y*, argument carries the superscript *ini*, standing for 'indefinite null instantiation'. Fillmore 1986 discusses the fact that some arguments of predicators can, when optionally omitted, be construed in a manner analogous to the existential closure of indefinites (Heim 1982, and much subsequent research; recently Ruppenhofer and Michaelis 2014).[24] The SEM value of (32) shows TDW *have* to contribute nothing to the semantics of a phrase in which it occurs – aside from the tense or aspect information it may host – despite the important syntactic role *have* often plays in the idiom.

The VP *have to do with Paris* illustrates Case 2. The head daughter of the construct shown in (35) contains a TDW *have* word based on the TDW *have* lexeme that has undergone the Indefinite Null Instantiation derivational construction and therefore lacks a valent corresponding to the *Y* valent in (32) and also contains an *exist-fr* (**5**) whose value *w* binds the DEGREE value in the *i-do-[rel]-fr* of the non-head daughter and the mother, as determined by (32).

[24] Indefinite null instantiation is implemented in SBCG by a derivational construction not shown here whose unique daughter contains a valent XP_x^{ini} and in which (i) the mother's VAL list lacks XP_x^{ini} and (ii) the mother's FRAMES list adds to the FRAMES list of the daughter an *exist-fr* whose bound variable (BV) value is *x*. Analogously, the Definite Null Instantiation Construction eliminates a daughter's XP_x^{dni} valent and inserts a *the-frame* in the mother's FRAMES list, binding *x*.

(35)



6.5 The TDW phrase as infinitival relative clause

We consider next an example of Case 3, *something to do with Paris*, as a freestanding NP in a sentence like (36)a. In the freestanding NP case there is no restriction to XPs expressing a measure of degree, as mentioned above and shown in (36)b.

(36) a. Every time I read something to do with Paris I end up crying.
b. I recently read an article to do with mares and foals.

We have noted that the string *something to do with Paris* has a different analysis in an expression like *The book has something to do with Paris* and an expression like *The book concerns something to do with Paris*. In the former case *something to do with Paris* is a VP in which *something* plays an adverbial role, like the word *plenty* in (37).

(37) The war had nothing to do with him. But it had plenty to do with Magdalena.

In the latter case *something to do with Paris is a noun phrase* and *to do with Paris* is an infintival relative clause modifying the nominal (NP or N) *something*. Whereas in the former case *to* is a verb, in the latter *to* is a complementizer. Unlike the infinitive verb TDW *to* the complementizer *to* that introduces TDW infinitival relatives is not a special, idiomatic word, restricted to TDW complements, as shown in (38).

(38)

complementizer-word	
FORM	⟨*to*⟩
ARG-ST	⟨*Y*:VP[VF *base*]⟩
SYN	CAT [LID ⟨*i-to[null]-fr*⟩, XARG *none*, SELECT [CAT *noun*], VF *inf*]; VAL ⟨*Y*⟩
SEM	INDEX *none*; FRAMES ⟨ ⟩

The listeme pictured in (38) differs from that shown in (30) by indicating that it is a *complementizer-word*, specifying a single argument: a VP complement constrained only to [VF *base*] (rather than one constrained to [LID <*i-do[rel]-fr*>]), by eliminating the XARG of (30), and by specifying a non-null SELECT value.[25]

We assume that words like *something, everything, somebody,* etc. belong to a type *quantifier-noun-word,* which is a subtype of both *noun* and *quantifier*. The listeme for the word *something* is given in (39).

[25] Nouns, N-bars, and NPs are all of category *noun*. We avoid taking a position on whether infinitival relatives modify N-bars or NPs.

(39)
$$\begin{bmatrix} \textit{quant-nominal} & \\ \text{FORM} & \langle \textit{something} \rangle \\ \text{ARG-ST} & \langle\ \rangle \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \textit{noun} \\ \text{LID}\ L \end{bmatrix} \\ \text{VAL} & \langle\ \rangle \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & x \\ \text{FRAMES} & L{:}\left\langle \begin{bmatrix} \textit{exist-fr} & \\ \text{BV} & x \\ \text{REST} & l \end{bmatrix}, \begin{bmatrix} \textit{thing-fr} & \\ \text{LABEL} & l \\ \text{INST} & x \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

We assume that listemes in this class contain both a *quantifier-fr*, e.g. *exist-fr,* and a RESTRICTION (REST) frame, viz. *thing-fr* or *person-fr*, although nothing here turns on that assumption.[26] Additionally we assume that at least some quantifier frames, including this one, are subtypes of *degree-quantifier-fr*. The NP *something to do with Paris*, is licensed by *something* (39), the infinitival relative clause *to do with Paris*, and the Head Functor Construction (40).[27] The derivation tree of the NP *something to do with Paris* as it appears in (36)a is shown in (41).

(40) **Head-Functor Construction** (↑ *headed-construct*)

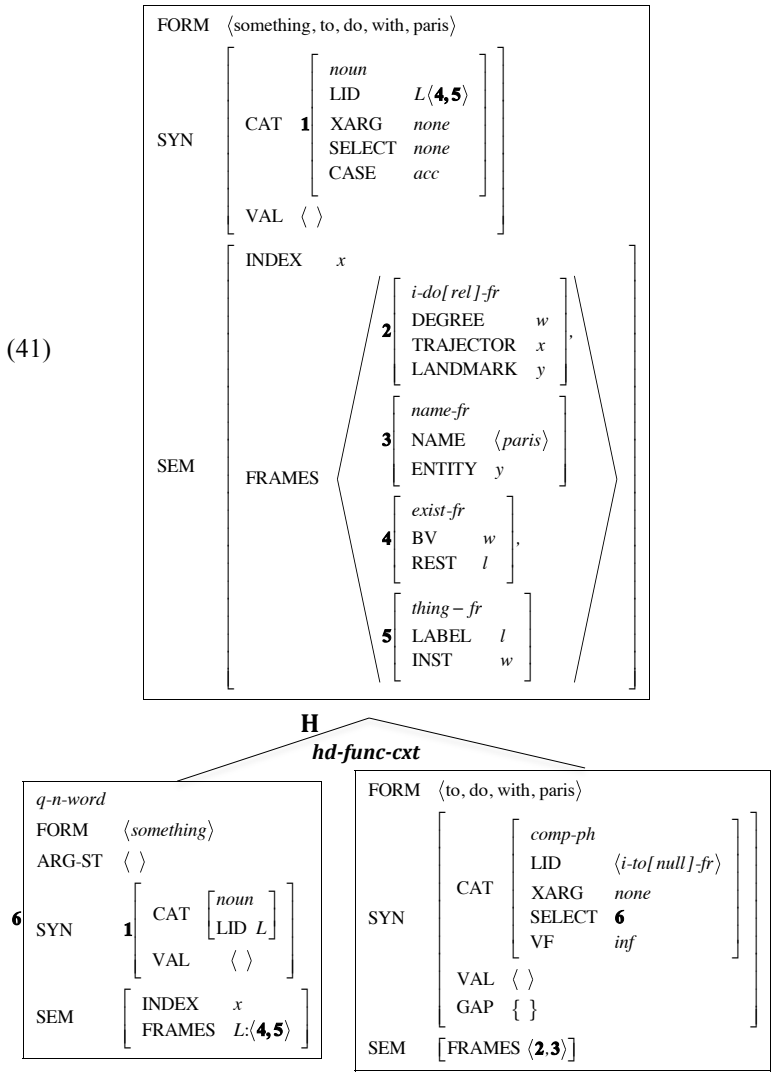
$$\textit{hd-func-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \left[\text{SYN}\ X!\left[\text{MRKG}\ M\right]\right] \\ \text{DTRS} & \left\langle \left[\text{SYN}\ \begin{bmatrix} \text{CAT} & \left[\text{SELECT}\ Y\right] \\ \text{MRKG} & M \end{bmatrix}\right], \right. \\ & \left. Y{:}\left[\text{SYN}\ X\right] \right\rangle \\ \text{HD-DTR} & Y \end{bmatrix}$$

[26] The LABEL feature of Minimal Recursion Semantics, of which we make only occasional use in this chapter, allows the representation of hierarchical relations among frames in the 'flat', list structures of MRS. In (39) The two tokens of the LABEL variable *l* show that *thing-fr* is the value of the RESTRICTION feature of *exist-fr*.

27 We assume for simplicity of exposition that all modifier-head structures are licensed by the Head-Functor Construction.

The Head-Functor Construction, based on the work of Van Eynde 2006, 2007 and Allegranza 1986, 1998, unites specifiers and modifiers as 'functors', bearing non-null values for the SELECT feature. The entire syntax of the mother is inherited from the head daughter except for the MARKING value, which is mentioned in (40) for completeness although it plays no role in this chapter.

(41)



In the head-functor construct *something to do with Paris* (41), the mother is a noun phrase, i.e. of type *noun* with [VAL < >], and inherits its CAT value from the head daughter *something.* The mother's XARG and SELECT features both have value *none.* In the mother's SEM value the INDEX *x* is inherited from the head daughter *something*. The *exist-fr* (**4**) and the *thing-fr* (**5)** are also inherited from the head daughter and the remaining frames, *i-do[rel]-fr* (**2**) and *name-fr* (**3**) are inherited from the non-head (modifier) daughter *to do with Paris.* The non-head daughter's CAT feature shows it to be *verbal*, with [LID <*i-to[null]-fr*>, [XARG *none*], [VF *inf*], and to SELECT the head daughter *something* (**6**). The VAL list is empty.

As we have noted, the constituent bearing [FORM <to, do, with, Paris>] in (41) is different from the constituent with the same FORM value in (31) and (35). The former is a clause, specifically an infinitival relative clause, while the latter is a VP. In the VP of (31) and (35) the trajector argument is successively raised from the XARG of *do* to the XARG of the auxiliary verb *to.* For the infinitival relative clause of (41), we follow the main lines of the analysis of simple infinitival relative clauses of Sag (1997: 469-470). The trajector XARG of *do*, instead of raising, appears in the GAP (née SLASH) set of *do* and does not appear on the VAL list, where it could be the target of raising. In this case *to* is not an auxiliary verb but a *complementizer.* The GAP content is constructionally bound off below the CP level, so at the clausal level, where the SELECT (née MOD) value is coindexed with the erstwhile solitary member of the GAP set, the GAP set itself is empty.[28]

7 Conclusion

We have shown that the idiomatic character of TDW, like that of many other multiword expressions, is exclusively lexical: both *to* and *do* are verbal heads with idiosyncratic valence requirements. Simply put, *to* selects a VP headed by *do* and *do* selects a PP headed by *with*. The resulting combination, licensed by the Predicational Head-Complement Construction, is selectable by an idiomatic subject-raising lexeme, *have* (or *be*), which also contributes a degree argument or licenses its existential closure.

[28] As discussed in footnote 15, for Sag (1997) the SELECT (MOD) value is an N' rather than an NP. Also the features have different names and in some cases somewhat different definitions, but are sufficiently alike for present purposes: GAP replaces SLASH, SELECT replaces MOD, [VAL < >] replaces [SUBJ <*PRO*>] in this context, and CAT replaces HEAD.

Alternatively a TDW phrase can constitute an infinitival relative clause, in which case there is no restriction of the preceding phrase to a degree-quantifier. Idiomatic *do*, despite its restricted combinatorial potential, is compositionally interpreted, denoting an association between two entities, the first of which is expressed by the non-locally-instantiated subject and the second of which is expressed by the object of the *with*-headed PP. SBCG has provided us with a precise framework in which to express this analysis and to predict exactly those syntactic configurations in which TDW occurs and does not occur.

Idioms frequently contain fewer words than is often supposed, e.g., by the compilers of idiom dictionaries and other presumed authorities. However, not all words that are necessary parts of an idiom occur only in that idiom. An example is the requirement of canonical case-marking *with* in the TDW idiom. The reverse is the case with TDW *have*, which illustrates the fact that some words which appear only in an idiom are not required in all instances of that idiom. For most idioms, the phrase-structural configurations in which their words can appear derive exclusively from the syntactic potentials of the words themselves, which often mirror the syntactic properties of canonical words with similar meanings, subject of course to idiosyncratic limitations. The syntactic privileges of occurrence of the *beans* of *spill the beans* is a subset of the syntactic privileges of occurrence of the word *secrets*. The meanings of the phrases and sentences in which most idioms occur are composed by the same processes as compose the meanings of phrases and sentences that contain no idiom words, and most phrasal idioms, properly analyzed, contain no phrasal information.

References

Allegranza, Valerio (1998) Determination and quantification. In *Linguistic Specifications for Typed Feature Structure Formalisms*, vol. 10 of *Studies in Machine Translation and Natural Language Processing*. Luxembourg: Official Publications of the European Communities. 281–314

Allegranza, Valerio (2007) The signs of determination: Constraint-based modelling across languages, vol. 16 of SABEST: Saarbrücker Beiträge zur Sprach- und Translationswissenschaft. Frankfurt am Main: Peter Lang.

Di Sciullo, Anna Marie and Edwin Williams (1987) *On the Definition of Word*. Cambridge, MA: MIT Press.

Kay, Paul and Ivan A. Sag (2012) Discontinuous Dependencies and Complex Determiners. In H. Boas and I. Sag, (eds.), *Sign-Based Construction Grammar*. Stanford: CSLI Publications. 211– 238.

Kay, Paul, Ivan A. Sag and Dan Flickinger (submitted). A lexical theory of phrasal idioms.

Fillmore, Charles J. (1986) Pragmatically controlled zero anaphora. *Proceedings of the Berkeley Linguistics Society* 12: 95 – 107.

Fillmore, Charles J. and Paul Kay (1995) *Construction Grammar*. Unpublished ms., University of California, Berkeley.

Gazdar, Gerald, Geoffrey Pullum and Ivan A. Sag (1982) Auxiliaries and related phenomena in a restrictive theory of grammar. *Language*, 591– 638.

Heim, Irene (1982) *The semantics of definite and indefinite noun Phrases*. PhD dissertation, University of Massachusetts, Amherst, Published in 1989 by Garland, New York.

Karttunen, Lauri (2013). You will be lucky to break even. In T. Holloway King and V. dePaiva, (eds.),, *From Quirky Case to Representing Space: Papers in Honor of Annie Zaenen*. Stanford: CSLI Publications: 167 – 180.

Michaelis, Laura A. (2012) Making the Case for Construction Grammar. In H. Boas and I. Sag, (eds.), *Sign-Based Construction Grammar*. Stanford: CSLI Publications. 31 – 69.

Nunberg, Geoffrey, Ivan A. Sag and Thomas Wasow (1994) Idioms. *Language* 70: 491 – 538.

O'Grady, William (1998) The syntax of idioms. *Natural Language and Linguistic Theory* 16: 279 – 312.

Pullum, Geoffrey K. (1981) The category status of infinitival *to*. *University of Washington Working Papers in Linguistics* 6, 55 –72

Ramisch, Carlos (2015) Multiword expressions acquisition: A generic and open framework, *Theory and Applications of Natural Language Processing* series XIV. Berlin: Springer.

Ruppenhofer, Josef and Laura A. Michaelis (2014) Frames and the interpretation of omitted arguments in English. In S. Katz Bourns and L. Myers, (eds.), *Linguistic Perspectives on Structure and Context: Studies in Honor of Knud Lambrecht*. Amsterdam: Benjamins. 57 – 86.

Sag, Ivan A. (2012) Sign-Based Construction Grammar: An informal synopsis, in H. C. Boas and I. A. Sag, (eds), *Sign-Based Construction Grammar.* Stanford: CSLI Publications). 69 – 202.

Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake and Dan Flickinger (2002) Multiword expressions: A pain in the neck for NLP. *Lecture Notes in Computer Science* 2276: 1 – 15.

Sag, Ivan A., Thomas Wasow, and Emily M. Bender. 2003. *Syntactic Theory: A Formal Introduction*. Stanford: CSLI Publications, 2nd ed.

Van Eynde, Frank (2006) NP-internal agreement and the structure of the noun phrase. *Journal of Linguistics* 42:139 – 186.

Van Eynde, Frank (2007) The big mess construction. In S. Müller, (ed.), *The Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar*, Stanford University. Stanford: CSLI Publications. 415 – 433.