

Contextually determined fixity and flexibility in “thing” sentence matrixes¹

GERALD DELAHUNTY

Abstract

This paper describes the semantics and pragmatics of Thing sentences (TSs) and derives them from the sentence’s lexis and syntax. It describes several rhetorical and discourse management uses of the form and argues that these arise because the TS clause is simultaneously presupposed and focused. It shows that TSs are realized as a set of formal variants appearing to manifest a range of fixity and flexibility: some may be produced either analytically or holistically, others only holistically. Data from the Corpus of Contemporary American English (COCA) and other sources shows that most TSs occur at the informal end of the register spectrum, in registers typically produced under time pressure, which is also where the less flexible variants tend to occur. The paper proposes a source and conditions for the creation of discourse management expressions like TSs and identifies linguistic elements from which they may be constructed and a diachronic trajectory for their origination, development, and extinction.

Keywords: copula; corpus; COCA; discourse organization; fixedness; formulaicity; phraseology; register; specificational; shell nouns; subordination; thing; thing sentence.

1. Introduction

This paper investigates, from a phraseological point of view, the variant forms, discourse uses, and register distribution of a sentence type I call “*Thing* sentences” (TSs). It is a part of a broader research project that seeks to identify the linguistic resources that speakers of a

language use to create new linguistic devices for discourse organizational purposes, and to describe and explain the pragmatics of a number of the devices so created (Delahunty 2001, 2006, 2007, 2008, in prep). These devices include a range of non-canonical sentence types constructed from semantically minimal lexis such as the copula, pro-forms, “shell” nouns (Schmid 2000), and the subordination of the clause representing the primary information to be communicated by the sentence (Pusch 2006).

Thing sentences are a subtype of specificational copular sentence (Higgins 1976) whose subject head is *thing*, which is characteristically definite (Brenier and Michaelis 2005), and whose complement is a finite clause:

- (1) *the thing is that Black and White America doesn't exist* (COCA ACADEMIC)

TSs allow several formal variants that differ in their internal organization and are differentially associated with factors such as style, register, and mode (Delahunty 2007, 2008, in prep). I refer to the variant described and exemplified just above as the “full TS.” Other variants are created by omitting the definite article:

- (2) *Thing is that CDOEXM isn't supported from ASP.NET...* (<http://forums.asp.net/t/1219872.aspx>)

Or the conjunction:

- (3) *The thing is, it actually works* (<http://www.beastwithin.org/users/wwwolf/games/nwn/gametools/neverblender>)

Or both the article and conjunction:

- (4) *Thing is, I do that most of the time anyway* (<http://www.kevinleitch.co.uk/wp/?feed=rss2>)

Although I will argue that the internal organizing principles of these variants are different from those of the full TS, their meanings and pragmatic functions are identical, so I include them among the *Thing* sentences that are the focus of this paper. I refer to *(The) thing is (that)* as “the matrix,” and the tensed S as “the (complement) clause.”

Additionally, the copula may be doubled:

- (5) The thing *is is* that we haven't told John yet (Tuggy 1996: 713. See also Massam 1999; McConvell 1988).

This is the variant that has attracted most research attention, as scholars attempted to account for *is is*. Because Brenier and Michaelis (2005) offer a compelling optimality based account for the double copula, viz., it is a correction of the mismatch between the intonation contour and syntax of the sentence, I do not address that issue in this paper.

There are also sentence types that are lexically and syntactically related to TSs. One of these has *thing* as subject head but with pre-modifiers and/or post-modifiers, e.g., *The important thing is ...*; *The first thing is ...*; *The thing about X is* Because of these modifiers, the semantics and therefore the pragmatics and discourse functions of these sentence types differ from those of TSs, e.g., they may order a sequence of (sub)topics, and so are not included in this paper. (See Delahunty and Velazquez-Castillo 2002 for a discussion of these and other related sentence types).

The complement of these copular sentences may also vary. It may be an NP:

- (6) The thing is also money (<http://www.ungei.org/infobycountry/247-1214.html>)

Or a *to*-infinitive:

- (7) The thing is to not be on the receiving end of fate, ... (<http://www.hp1.hp.com/news/2002/jan-mar/ramani.html>)

Because these have different semantics and discourse potentials from TSs (e.g., the NP complement is not propositional and the *to*-infinitival forms have an exhortative function lacking in TSs), I do not deal with them here.²

Discourse uses of *thing* also occur in a range of other expression types, e.g., *OK. Here's the thing* (Restasis TV ad), and *The thing is this*: (Sirr 2009: 34). These, too, must await future research, though Schmid (2000) discusses a number of them.

Thing sentences are a subtype of specificational copular sentence which allows a potentially infinite range of subjects. The head of these subjects is

a noun whose semantic specificity may range from the very specific, such as *milagro/miracle*, in:

- (8) El *milagro* es que estos pinos se mantengan.
 The miracle is that these pines themselves maintain
 The *miracle* is that these pines survived.
 (Corpus Oral de Referencia del Español Contemporáneo)³

to the maximally general *thing* that we find in TSs. Like Schmid (2000), I believe that a crucial analytic distinction must be made between TSs and sentences with semantically richer subject heads: the noun in the non-TS type names a relationship, e.g., *implication*, between the complement clause and its context and thereby evokes a (cognitive) frame in which the relevant context and complement clause are related, whereas the semantics of *thing* is so non-specific that it cannot do this, with the consequence that the relationship between the TS clause and the context must be otherwise accounted for, as I demonstrate below. (See also Delahunty 2008 in prep; Delahunty and Velazquez-Castillo 2002; Francis 1994)

For the purposes of this paper, I have divided the TS variants into the five types in Table 1.

One goal of this paper is to show that the TS-type is situationally dependent. That is, it is an expression “with contextually restricted conditions of use” (Kuiper et al. 2007: 317: cited in Edmonds 2010a: 11) or is “appropriate to a situation of a certain kind which is appropriate relative to certain communicative ends” (Coulmas 1981: 16: cited in Edmonds 2010a: 24). Edmonds adds, “With respect to conventional expressions, it is predominantly

Table 1. *TS variants*

-
- a. Full TSs: *The thing is that S <the thing is S>*¹
 b. *The thing is* + punctuation marker (typically a comma) without *that <the thing is,>*²
 c. *The thing is* without *that* and without any punctuation marker *<the thing is>*
 d. *Thing is* + punctuation marker without *that <thing is,>*
 e. *Thing is* without any punctuation marker and without *that <thing is>*³
-

1 I use < > to demarcate a cited form from surrounding text, and especially to disambiguate a comma, thus distinguishing *<thing is,>* from *<thing is>.*

2 For brevity's sake and because the comma is the most frequently occurring punctuation mark in my data, I use “comma” to represent all the punctuation that occurs between the TS matrix and clause.

3 The paradigmatically related sentences with subject heads other than *thing* exhibit corresponding variants with similar internal organization.

the association with a social (e.g., classroom discourse) and/or pragmatic (e.g., different speech acts or *conversational management*) situation that is of interest" (2010a: 24; emphasis added). Edmonds also says, "Concretely, this implies that an identical surface string may be a formula in certain communication situations but not in others" (p. 13). As I show, the TS is a device with a specific discourse, and typically, conversational function. Its formal variants have the same characteristic interpretation and functions, though they differ in their internal organization and apparent degrees of fixity, and are associated with different registers and modes of production, though the full version appears to be capable of either holistic or analytic production, depending on the specifics of the situation, as allowed for by Edmonds and the research she cites.

Full TSs are syntactically organized according to the grammar of English while the punctuated variants are the result of the intonational separation of the matrix from the clause, and the ellipses result from very general reductive processes, presumably facilitated by the situational dependence of the TS-type. Additionally, the distribution of these variants across registers is such that the fuller variants occur in registers typically produced under little time pressure but which require considerable attention to language and prescriptive correctness, whereas the most reduced, most fixed variants occur in registers typically produced under considerable time pressure but which require little attention to language and prescriptive correctness.

2. Data sources

The data for this paper comes from several sources: Mark Davies' Corpus of Contemporary American English (COCA), the World Wide Web using WebCorp (Bergh 2005), five Inspector Rebus novels by Ian Rankin, and occasional items from whatever I happened to be reading. The data consists of instances of the sentences along with as much of their context as needed for adequate interpretation, or as much as the concordancers allow. COCA consists of five subcorpora: ACADEMIC, MAGAZINE, FICTION, NEWSPAPER, and SPOKEN texts.⁴

It's important to be aware of the types of texts in each subcorpus. ACADEMIC includes texts from publications such as *American Scholar*, *Africa Today*, *World Affairs*, *Anthropology Quarterly*, *Style*, and many others, representing a range of texts, some more highly edited than others. Some include representations of speech.

FICTION includes texts from books and from magazines such as *The Atlantic Monthly* and *Literary Review*. Fiction, of course, characteristically

includes representations of speech, typically intended to be a realistic if conventionalized representation of conversation or of interior monolog.

MAGAZINES includes texts from publications such as *Money*, *American Heritage*, *Rolling Stone*, *Prevention*, and *Newsweek*. This latter magazine has adopted an informal style to survive challenges from electronic news magazines.

NEWSPAPERS includes texts from *The New York Times*, *Associated Press*, *USA Today*, and the *San Francisco Chronicle*, representing a range of styles and including representations of speech.

SPOKEN consists of transcriptions of talk on TV news programs, including ABC 20/20, ABC Primetime, and CNN Crossfire. This register is rather different from casual conversation amongst friends. The speakers are professional newscasters, journalists, and commentators, so the linguistic style is often of texts written to be spoken, and as journalists are often conservative in their language attitudes, the style tends to be prescriptively normative.

3. Uses and rhetorical strategies

In this section, I briefly describe how TSs are used in discourse and the rhetorical strategies they instantiate. I also explain why TSs function in these ways by showing how their functions can be derived from their lexico-grammatical characteristics interacting with their contexts and general pragmatic principles, without recourse to construction-specific stipulations. I will then return to the main topics of the paper: the register distribution of the TS variants and an explanation of that distribution. (For more detailed argumentation, see Delahunty in prep.)

3.1. *What TSs mean*

As the TS is a sub-type of specificational sentence, it has a definite subject which represents a variable whose value is represented by the clause (Higgins 1976). In other words, the subject represents a description that the complement clause instantiates, or the subject characterizes the clause (Schmid 2000). Unlike semantically richer nouns, *thing* denotes merely COUNTABLE ENTITY and so does not specify how the proposition represented by the complement is to be related to its context (Delahunty and Velazquez-Castillo 2002; Schmid 2000).⁵ As TS subjects are also characteristically singular (see Section 5), only a single value or instantiation is relevant.

Because the TS subject is definite, it licenses the presupposition that an identifiable entity exists, but as (*the*) *thing* has no situational referent, it is

interpreted as the speaker's assumption that the hearer accepts the description it denotes and that a contextually relevant entity exists that satisfies that description, namely, the proposition denoted by the complement. Consequently, that proposition is backgrounded rather than asserted.

However, in addition to being presupposed, the complement clause is doubly focused, first, by virtue of its predicate position, which Lambrecht (1994) characterizes as the default focus position and, second, because it represents the value of the variable represented by the TS subject, it is an argument, which Lambrecht claims is a marked focus. The complement clause is thus both presupposed and markedly focused and thus introduced into the discourse as a new and especially significant member of the set of propositions to be taken for granted in the local discourse.

These apparently contradictory interpretations are used to indicate that the speaker is proposing that the audience take for granted the proposition represented by the TS clause and as the ground for subsequent utterances, replacing the proposition represented by an utterance prior to the TS which would otherwise have functioned as that ground. For example:

- (9) When the barman slouched back with Rebus's change, Hogan greeted him by name.
"Okay, Malky?"
The young man frowned. "Do I know you?"
Hogan shrugged. "Thing is, *I know you*." He paused. "Still on the smack?" (Rankin 2003: 192)

Malky's "Do I know you?" should have been the basis for the next contribution to the discourse, e.g., a yes/no answer. Hogan's TS preempts this topical trajectory and substitutes his "*I know you*" as the basis for further conversation, as his "Still on the smack?" demonstrates. (Delahunty 2007, 2008, in prep provide more detailed arguments.)

3.2. *Uses and rhetorical effects*

The semantics and pragmatics of TSs allows them to be used for several rhetorical effects, of which I describe just three here (see Delahunty in prep for others). TSs may contradict a prior assumption:

- (10) I imagine that the hour you are onstage each week is high-energy and exhausting. Do you have fun doing the show each week?

What's the funniest unexpected thing that's happened onstage during the run?...

The thing is that since the plays are done in random order and, since there is a lot of audience participation, the whole thing is rather unexpected. (Too Much Light Makes the Baby Go Blind)

In this example, the assumption – a presupposition presented as an element to be taken as common ground by the interlocutors – is that there exists a unique “funniest unexpected thing that happened onstage during the run,” licensed by the definite article and superlative “funniest.” This presupposition is implicitly contradicted and replaced by the proposition represented by the TS in the interviewee's reply. This is consistent with Carter and MacCarthy's (2006) contention that TSs represent problems and with Tuggy's (1996) contention that they represent “disconformity” with their contexts, a characterization endorsed by Schmid (2000).

TSs also affect topical development. Aijmer (2007: 43) claims that a TS “signals a new” topic or subtopic. However, none of my examples effects a shift from one topic to another, while all of them signal a shift from one subtopic to another, as illustrated by:

- (11) just got this new 2003 voyager. *The thing is that, the other day it didn't start in the morning* (would not turn the engine). I try it again and would you believe it? No lights come on on the dashboard, won't turn the engine nothing is happening. Change the battery, check the fuses, etc etc. What is it with it? Could you help please? Ta very much. (<http://www.faqs.org/qa/qa-3417.html>)

This is from a website to which one writes about cars. The first sentence sets up minimal expectations about how the topic will develop. The TS marks the subtopical shift from the fact that the writer had just got a new Voyager to the (unexpected) fact that it wouldn't start. The TS clause provides very little information, and thereby makes relevant further information, in this case, a description of the car's symptoms, a list of things the owner did to remedy the problem, and a request for help. In fact, TSs often function as topic sentences – projecting the content and point of view of what follows them.

Hudson (1998: 143), citing Finell (1996), suggests that topic changing discourse markers also “soften potentially impolite changes.” While I don't believe that TSs are discourse markers (yet), the TS in (10) politely shifts subtopic. Politeness theory (Brown and Levinson 1987) claims that redressive action frequently requires elaboration of the message. As the proposition

represented by a TS clause is the sentence's message, we can view the matrix as elaboration which may indicate increased politeness. Politeness theory also claims that impersonalizing a message may indicate increased politeness. Because TSs represent their propositions as presupposed rather than asserted by the speaker, they are politely impersonal. All of these features can be seen in (10) above.

A TS may also function as a "pre-" (Terasaki [1976] 2004), that is, as an expression that prefigures an upcoming utterance and bids for the interactional space it needs. TSs are often informationally less than optimally relevant in their contexts but in such a way as to provide audiences with clues to the information the speaker would provide if granted the opportunity. For example:

- (12) Mackenzie was shaking her head. "These records are up-to-date. The last rent money we received was only last week. It was paid by Mr. Baird."

"You're thinking he sublet?"

A broad smile lightened Mrs. Mackenzie's face. "Which is strictly forbidden by the tenancy agreement," she said.

"But people do it?"

"Of course they do. *The thing is, I decided to do some sleuthing myself ...*" She sounded pleased with herself. Rebus leaned forward in his chair, warming to her.

"Do tell," he said.

"I checked with the city's other housing areas. There are several Robert Bairds on the list. Plus other forenames, all with the surname Baird." (Rankin 2005: 84)

Mackenzie uses her TS to bid for an extended next turn. The TS anticipates, "projects," to use Hopper & Thompson's (2008: 105–6) term, further talk by its producer, and Rebus' "Do tell" indicates that he accepts her bid and cedes interactional space for additional talk. This kind of topical projection is a characteristic rhetorical function of TSs.

4. Fixity/formulaicity

In this section, I examine TSs against a set of criteria that have been used by researchers to determine whether expressions are fixed or formulaic.

Edmonds (2010a, 2010b) provides a valuable review of the literature on fixity and distinguishes two major analytic points of view – the

conventional and the psycholinguistic, though she recognizes that many researchers invoke both points of view. The former views fixed phrases as “sequences with a stable form that are used frequently by speakers in certain prescribed social situations,” (Bardovi-Harlig 2009: 757). The latter views fixed phrases as “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at time of use, rather than being subject to generation or analysis by the language grammar” (Wray 2002: 9). The data show that TS matrixes have a stable form with some variation and a specific meaning and pragmatics and are used for specific discourse management purposes. Though my data cannot address how TSs are stored, it is consistent with holistic storage of all variants and analytic generation of one in certain contexts.

The remainder of this section discusses researchers’ claims that expressions consisting of more than one word may be fixed if speakers intuitively deem them so, if they are situationally dependent, in community-wide use, occur frequently, are semantically opaque, non-compositional, syntactically incoherent, less costly to produce, more fluently produced, invariable, of greater complexity than a speaker seems otherwise capable of, and extend the uses of the expression beyond those of adult or native speakers. These last two criteria are irrelevant here as my data cannot address them.

Full TSs certainly consist of more than a single word, though the <*thing is*,> variants, typically written as two words, may be produced as a single intonational unit, though not with word stress.

My intuition suggests that the TS variants represent a cline of fixity, with full TSs least fixed and <*thing is*,> most fixed; that is, the full TS variant may, on occasion, be generated analytically and on other occasions may be produced holistically, while the <*thing is*,> variant is always produced holistically. Wray (2002) is properly skeptical of the reliability of intuition, primarily because corpus findings show that we have poor intuitions about statistical tendencies.

The clearest examples of situationally dependent expressions may have no words in common, e.g. limericks and expressions of the class represented by *Is the pope Catholic?* (Moon 1998: 95). About these, Wray says, “With the absence of common lexical or morphological forms across the set members, the formulaicity resides in the context-structure interface rather than the form per se. This challenges the idea of a form-based criterion, for what, precisely, is being stored when all the words can be novel?” (Wray 2002: 32). Applying this to TSs, we might say that they are one of a number, whose size we cannot determine, of ways in which a speaker may

communicate or indicate disagreement with some proposition in the relevant prior discourse along with the intention to propose an alternative proposition as the common ground for the continuation of the discourse. For example, to "Can I have this transferred to my phone?" one might answer, "Well, the connection's not working," or "What on earth for?" or "The thing is that it's not my phone you're calling from." (modified from London-Lund Corpus). That is, at the point at which a TS occurs, the normal requirements of the interaction require a contribution to the discourse with certain characteristics, and the TS is one of indefinitely many ways in which these requirements may be met. That the speaker chooses a TS rather than one of the alternatives suggests that the TS is the best choice among those alternatives for their purposes at that point, or it was the least costly form to produce, and the least costly form is acceptable in the context and is compatible with the speaker's "abilities and preferences" (Wilson & Sperber 2004: 612). This suggests that TSs, as a type, are situationally dependent; that is, regardless of whether they are formally fixed, they perform a specific pragmatic function, viz., indicating that the complement clause is both presupposed and marked, with consequential discourse effects.

TSs are in community wide use – the expression is known to all or most of the members of the speech community. The variants occur in a broad range of COCA texts, in Scottish fiction, in the Wellington Corpus of New Zealand English, in blogs and other online texts, and native speakers, at least as represented by the US university students I have discussed them with, recognize the form and can roughly characterize how they use it.

On the basis of their frequency, Biber et al. (1999: 992) identify lexical bundles as "word combinations" that occur "at least ten times per million words in a register" and "must be spread across at least five different texts in the register." They identify *but the thing is* as a four-word bundle that occurs in conversation, though they are silent on the other variants, presumably because these did not meet their frequency requirements. My data collection techniques do not allow me to determine how frequently TSs occur per some benchmark number of words, though they do allow me to say how many actual TSs, and which variants, occurred in the first 100 *thing is* hits in each of the five registers in COCA (see Table 2 below).⁷

Wray (2002) is skeptical about the reliability of frequency counts in determining fixity, as these may be undercut by intuition-based decisions about which expressions are worth counting, the size of the corpus, and the difficulty of determining the boundaries of the target expression. And we do find that expressions that are clearly fixed, e.g., semantically opaque

idioms, occur very infrequently (Moon 1998: 79), so clearly, frequency and fixity are logically separate. And because holistic storage is a characteristic of an individual idiolect and corpora are (typically) composed of texts produced by many individuals, frequent corpus occurrence of an expression can only be suggestive of its fixity; other criteria must be invoked to demonstrate it. However, if we took Wray's skepticism about frequency to its extreme, we would have to abandon a useful heuristic for identifying potentially fixed expressions. In any case, a search for *thing is* yielded only 50 TSs out of the first 100 hits for each of COCA's textual categories, while in the five Rankin novels (approx. 2000 pages), I found only 24, and my WebCorp search yielded only 36.

Regarding the compositionality of fixed expressions, Wray remarks (2002: 33) that, "[a]t the heart of this approach is the observation that a sequence of words, once it is formulaic, is subject to detachment from the effects of the live grammar and lexicon. The string is no longer required to be grammatically regular or semantically logical. Sequences become frozen, or fossilized, and as a result often retain words or grammatical forms that are no longer current in the language." To apply this to TSs, we must ask, "What is included in the live grammar and lexicon?" As we have seen, full TSs are semantically transparent, compositional, and grammatically coherent, and so need not be fixed. However, a COCA search for the grammatical variants *thing was*, *thing will be*, *thing is not*, *things are*, and *things were* returned few or no instances (see Section 5) indicating that speakers do not make use of the full range of grammatical alternatives made available by the grammar, suggesting that they are not "unrestrictedly constructed" (Mel'čuk 1998: 26).⁸

Additionally, one might question whether variants without *the* are transparent, and whether variants without *the* or *that* are grammatically coherent. The determination depends upon whether processes such as the omission of *the* and *that* are part of the "live grammar" of the language. Clearly, reductions may be the result of very general processes. *That* deletion is not limited to TSs and is generally regarded as a grammatical rule (e.g., as a "structural matter" by Quirk et al. 1985: 900; see also Huddleston and Pullum 2002: 949–954). According to Biber et al. (1998: 681), *that* omission is rare in academic prose, most frequent in news, and favored by the verbs *think* and *say*, which are like *thing* in being semantically quite non-specific. Dor (2005) claims that *that* deletion indicates increased speaker commitment to the truth of the proposition represented by the clause, a formulation that is consistent with the sense of presupposition associated with TS complements that I showed in Section 3.1.

The deletion is also not limited to TSs. In the Rankin data I found several other head nouns without *the*, e.g., *word is*, *fact is*, *bugger is*, *trouble is*, *story is*, as well as the plural *chances are*, functioning as the subjects of specificational sentences. Quirk et al. (1985) treat *the* deletion primarily as a "grammatical feature" (p. 1511 Note [b]), but also as "situational ellipsis" (p. 899), whereby words "that normally occur before the onset of a tonic unit ..., and hence have weak stress and low pitch" may be omitted by "reductive process[es] on the phonological, rather than on the grammatical level" (p. 896, esp. Note [b]). According to Quirk et al. (1985: 883–890), reductions such as those displayed by TSs must be recoverable verbatim, a characteristic which would be facilitated by the situational dependence of TSs. (See also Biber et al. 1999: sections 3.7.5 and 14.3.5.1.) Carter and McCarthy (2006: 186–187) merely say that "articles considered obligatory in formal speech and especially in writing may be unnecessary in informal speech when the referent is obvious." To informal speech we can add informal writing, or perhaps more accurately, writing that indexes aspects of speech, especially its reduced need for explicitness. (Huddleston and Pullum 2002 seem not to deal with this issue.)

Regardless of how we understand these elisions, the elements of TSs that may be delimited by punctuation and which seem to function as units relative to the complement, <*the thing is that*,>, <*the thing is*,>, <*thing is*,> are not phrases and therefore not generatable by the grammar, suggesting that they may be fixed by Wray's criterion.

Written data is far from ideal in addressing whether TSs are produced with greater fluency and speed and with less internal disruption by restarts and hesitation phenomena than non-fixed expressions. Nonetheless, TS matrixes, especially the reduced ones delimited by punctuation, seem to fit the criterion of uninterruptability.

Wray (2002: 34) says that very few fixed items are "entirely fixed," and researchers allow for some variation in otherwise fixed expressions, including "open slots." While the specificational relationship between the subject and the complement is invariant, we can represent the TS variants with the formula: *(the) thing is/was (that) S*.

However, a formula of this kind is inadequate as a representation of the TS variants: it implies that the elements are organized only as a linear sequence and that the variants are related to each other merely by changes in tense and the optional omission of *the* and/or *that*, when in fact, the variants represent at least two different internal organizing principles.

According to Wray (2002: 37), Jespersen ([1924] 1976: 83), "sees stress as a key defining feature of formulaic sequences. Because a sequence is

‘felt and handled as a unit,’ he argues, it is not possible to pause between the component words or to stress them in an unaccustomed way.” While I have seen no evidence that *<thing is,>* may be interrupted, the punctuation that may occur in all TS variants appears to be the written indication of the intonational separation of the TS matrix from its complement clause, as well as tonic stress and intonation on *is*.

Because they may be fully syntactic and compositional, the full TS forms should be most likely to occur in registers that allow for and expect considerable attention to language, manifested in greater explicitness, while the other variants should tend to occur in registers expecting less attention to language and less explicitness. This is consistent with Halliday’s (1987) claim that written language tends to be syntactically organized, whereas spoken language tends to be intonationally organized. As the intonational TS variants occur in written registers as well as spoken ones, they are typical of speech – and speech-like – registers.

Jespersen’s observation and my analysis are also consistent with Hopper & Thompson’s (2008: 99) claim that in English *wh*-cleft and extrapositive sentences, “[p]rosodically, each clause typically forms its own prosodic unit, the first ending with a ‘continuing’ terminal pitch contour.” In this context “clause” refers to expressions which in formal spoken and written texts would be “reformulated” as a complete clause but are incompletely clausal in speech.

We have arrived at the tentative conclusion that TSs may be fixed at two levels. The first level is that the expression type is a situationally dependent specificational copular sentence; the second is that the TS variants represent the operation of contextually determined reductions and reorganization, presumably facilitated by the situational dependence of the expression type, and manifesting a cline of fixity from full TSs to *<thing is,>*. I turn next to a discussion of the distribution of the TS variants across registers.

5. Register distribution of TS variants

From a Hallidayan (1987) point of view, the TS variants, with their vague lexis, intonational rather than syntactic organization, and rightward placement of important information should occur primarily in speech(-like) registers. We should also expect the TS variants, ranging from full forms *<the thing is that S>* via increasing reductions and intonalization to the most reduced form *<thing is,>*, to correlate with registers ranging from those composed under least time pressure and in which the greatest degree of surface

correctness and the broadest lexical range are expected, through registers composed under increasing time pressure with decreasing expectations of surface correctness and lexical diversity, to those registers composed under the greatest degree of time pressure and with the lowest expectation of surface correctness and lexical range. And this is in fact what we find.

I searched COCA for instances of the strings *thing is*, *thing is not*, *thing was*, *thing will be*, *things are*, *things were*. All of these searches returned instances which upon investigation were not TSs, though some returned actual TSs. I identified all the TSs in the first hundred hits from each sub-corpus, and categorized these according to the variants distinguished in Table 1, and compared the numbers of variants returned in each register.

The searches for *things are*, *things were*, *thing will be*, and *thing is not* returned no TSs in the first 100 hits. The notable non-occurrence of these TS variants indicates that speakers do not make use of the range of variants made possible by English grammar and lexis. The non-occurrence of the plural variants may be due to the uniqueness implicature associated with focusing, while the non-occurrence of the negative variants may be due to the fact that a TS complement is focused and typically new information which replaces an assumption derived from the prior context and which functions as the basis for following text, and so would appear to be incompatible with negation.

The search for *thing was* returned 2934 hits, but only five of the first 100 were TSs; none of these were full TSs, four were <*the thing was*>, all from fictional narrative, and one was <*thing was*>, from a first person fictional narrative. The infrequency of past tense TSs – merely 10% of the number of present tense TSs – may be due to the two factors. First, because of the semantic vagueness of their matrix clauses, TSs are more likely to occur in speech(-like) contexts, especially in conversation, in which the present tense predominates (see Figure 6.1 in Biber et al. 1999: 456). Second, TSs comment on an aspect of their contexts, a meta-communicative act, which is most likely to occur in speech(-like) registers and especially in conversation, which Biber et al. (1999: 1045–1047) characterize as “interactive,” and thus to favor the present tense. The infrequency of past tense forms and the non-occurrence of the other possibilities may be taken to indicate that TSs are grammatically limited, and thus evidence of fixity, though we should also take into account that the forms might simply be incompatible with the discourse functions of TSs.

Because the past tense variants are so infrequent and their distribution is so entirely consistent with that of the corresponding present tense ones, it provides no information not available from the distribution of the present

tense variants, and so I omit it from the discussion the register distribution of TS variants.

A COCA search for *thing is* yielded only 50 in the first 100 hits. Table 2 shows how the variants distribute across the registers.

Table 2. *COCA distribution of TS variants*

	<i>FICT</i>	<i>SPOK</i>	<i>NEWS</i>	<i>MAGS</i>	<i>ACAD</i>	<i>TOTALS</i>
<i>Thing is</i> hits	1554	5319	1676	1540	473	10562
Actual TSs < 1st 100 hits	18	12	11	6	3	50
%age of TSs in register	36%	24%	22%	12%	6%	100%
Full TSs < <i>the thing is</i> that <i>S</i> >	5.6%	33%	9.1%	none	67%	14.5%
< <i>the thing is,</i> >	67%	50%	64%	100%	none	58%
< <i>the thing is</i> >	11%	16.7%	18.2%	none	33%	19.1%
< <i>thing is,</i> >	11%	none	9.1%	none	none	9.1%
< <i>thing is</i> >	5.6%	none	none	none	none	7.3%

The decline in the frequency of TSs from FICTION to ACADEMIC clearly shows a distribution skewed toward the casual, speech, and speech-like end of the register spectrum: there are six times as many TSs in FICTION as in ACADEMIC (18 vs. 3). The TOTALS column shows that reduced variants occur almost six times as often as the full variants (85.5% vs. 14.5%), and that the main reduced variant is <*the thing is,*> (58%).

In FICTION we find the broadest range of variants, of which only one 5.6% is full, while the vast majority (78%) are punctuated and therefore show signs of intonational organization, suggesting a degree of fixity.

SPOKEN had 12 TSs, of which 33% (4) were full, 50% (6) were punctuated, and none were <*thing is,*> or <*thing is*>.

We should expect NEWSPAPERS to share features with ACADEMIC and SPOKEN, given that COCA's SPOKEN is represented primarily by professional TV news personalities. We find that NEWSPAPERS and SPOKEN have very similar rates of TS occurrence: 22% and 24% respectively. In NEWSPAPERS, 9.1% (1) are full and 73.1% were punctuated, and none were <*thing is*>.⁹

In MAGAZINES we should expect a distribution like that in FICTION and NEWSPAPERS. However, 100% (6) of the TSs found in MAGAZINES are the <*the thing is,*> variant, and 83% (5) are in first person narratives and/or informal contexts.¹⁰

ACADEMIC texts had only 3 TSs; of these 2 (67%) were full TSs and one (33%) was <*the thing is,*>, which was transcribed from speech. In

another, more homogeneous academic corpus of one million words of business, civil engineering, and environmental sciences journal articles (Bond 2007), not a single TS occurred. Clearly, TSs occur very rarely in academic texts and we can tentatively conclude that when they do occur in those texts, they do so in their full form, unless there is a local reason not to.

The TSs collected from the Rankin novels and by using WebCorp (Bergh 2005) also display the kinds of reductions and skewed register distribution found in COCA. Table 3 shows that every Rankin TS is reduced and occurs in dialog.

All of the TSs found in the Rankin novels occur in dialog and none are full TSs. 25% (5) are <the thing is,> and 75% (15) are the further reduced <thing is,>. As all are punctuated, they show signs of intonational organization, suggestive of relative fixity.

The TS variants returned by a WebCorp search for the string *thing is* are presented in Table 4.

Table 3. *TSs in five Ian Rankin novels**

<i>Thing sentences</i>	20	
	DIALOG	OTHER
<the thing is that S>	none (0%)	none
<the thing is,>	5 (25%)	none
<the thing is>	none	none
<thing is,>	15 (75%)	none
<thing is>	none	none

* Four instance of <thing was> occurred in the Rankin novels; one occurred in dialog and the other three occurred in free indirect speech, which is "constructed dialogue" according to Tannen 1986 and "heteroglossic" according to Bakhtin 1981 [1953], both cited in Johnstone 2008: 60.

Table 4. *WebCorp examples**

<i>Thing sentences</i>	36
<the thing is that S>	8.3%
<the thing is,>	64%
<the thing is>	11.1%
<thing is,>	14%
<thing is>	2.8%

* The WebCorp search also returned two instances with doubled copula.

These are primarily from personal blogs and represent a very informal and highly involved style which include lots of contractions and elisions, as well as being minimally edited. It is notable, even remarkable, that 78% of these instances are punctuated, and therefore intonationalized, suggesting an oral substratum for this kind of writing.

From the data and discussion, we can derive three main observations. First, full TSs are typically unpunctuated, suggesting that they may, at least occasionally, be analytically generated with syntactic integration of matrix and clause.

Second, both full and reduced variants allow punctuation, though the latter far more frequently than the former, suggesting intonational organization of the matrix and hypotaxis of the clause (Halliday 1987).

Third, the non- or infrequent occurrences of past tense, plural, and negative variants suggest that TSs are not “unrestrictedly constructed” (Mel’čuk 1998: 26), though other, functional, accounts are available.

6. Discussion

Though full TSs seem to be syntactically organized and semantically compositional, they have specific pragmatic and discourse effects and so are situationally dependent. That several of their grammatically possible variants rarely or never occur suggests that TSs are fixed to some degree, a conclusion consistent with the intonational separation of the matrix from the clause, indicated by the frequent occurrence of punctuation. The situational dependence and fixity of the type allow for general patterns of reduction, which in turn are consistent with a register distribution that is skewed toward online, unplanned production, contexts that allow (or perhaps strongly prefer) forms that leave linguistic information implicit and are generally holistically produced.

TSs appear to be on a diachronic trajectory similar to that described by Hudson (1998). Her goal is to identify the trajectory of fixation and to uncover the forces that underlie it. She proposes that the trajectory is from “*ad hoc*” expressions (i.e., expressions typically produced in conformity with the lexico-grammar of the language) through fixed expressions to univerbations (see Chapter 6). She proposes that underlying this trajectory are the “fixing forces” (see Chapter 4): reduction in salience of the meaning, grammatical role, and cohesive relations of the expression.¹¹

When we apply Hudson’s criteria to TS subjects, we find that their meanings are non-salient: they have no clear reference, indeed no reference at all; even in full TSs, they lack some of the characteristics of subjects,

most notably, they are not agentive and seem to be limited to singular number; and they have no salient cohesive (e.g., antecedent/anaphor) relations with other phrases in their contexts (Hudson 1998: 109).

Hudson proposes that univertation results in words belonging to one or other of the classes recognized by the grammar of the language, primarily to the heterogeneous class of adverbs. At this point in their history, full TSs seem to be still produceable as "*ad hoc*" expressions, though the most reduced TS form, <*thing is*,>, appears to be fixed, perhaps close to the point of univertation, possibly as a sentential adverb. In this respect, it would be similar to the expressions analyzed by Hopper and Thompson (2008), though they do not go so far as to classify them as adverbs.

At this point in its development, the most reduced TS variant, <*thing is*, >, seems not to be independent of the full variant: the latter can readily be reconstructed from it. However, the occurrence of reduced forms in publications such as *Newsweek* and of full forms with punctuation between *that* and the remainder of the clause (see Example 11 and fn. 8) suggest that the reduced forms may be beginning to become independent of the full forms and ready to colonize the more formal registers which now strongly prefer the full, unpunctuated forms. If they continue on this trajectory, they will begin to allow a greater range of expression types as complement, just as Hopper and Thompson (2008) demonstrate is the case for *wh*-clefts and extrapositives. At that point, the relationship between "matrix" and complement will become proceduralized, and thereafter may lose some of its current semantic and pragmatic characteristics, perhaps the presupposed status of the complement, leaving the matrix as a generic markedness marker, competing with other such markers such as the inferential *It is (that) S* form (Delahunty 2001; Delahunty & Calude 2011; Pusch 2006).

We might think of TSs and their variants as providing contextualization cues. Heller (2001: 258) quotes Auer (1992: 4) to the effect that "contextualization ... comprises all the activities by participants which make relevant, maintain, revise, cancel, ... any aspect of the context which, in turn, is responsible for the interpretation of an utterance in its particular locus of occurrence." Thinking about TSs in this way allows us to identify the proceduralization of variants of sentence types like the TS as one source of such discourse management devices.

As I noted in Section 4, the Rankin data includes a number of specificational sentences with subject head nouns other than *thing*, e.g., *story*, which also perform discourse management tasks. This suggests that the specificational sentence type is a prolific source for expression types that can be used for discourse management functions.

Because TS matrixes are composed of elements which individually occur frequently and have minimal semantic specificity, they should come readily to mind and thus facilitate discourse planning. Indeed, it makes good sense to create discourse structuring devices that require minimal processing so that attention may be primarily directed to the propositions the devices mark.

These remarks suggest a trajectory like the following for the creation of discourse management devices such as TSs:

1. Speakers recognize a discourse need, e.g., to revise, cancel, or contrast an assumption derivable from the prior context, and create a form out of appropriate available linguistic elements to address that need.
2. Speakers recognize the effectiveness and particular “fit” between the new form and its discourse effects.
3. The form becomes established as a preferred one for its purposes.
4. Efficiencies set in and the form is subject to general patterns of intonationalization and reduction: it is realized as a range of variants that are differentially distributed across registers. (This is as far as TSs seem to have come.)
5. The form becomes independent of its linguistic origins and colonizes registers from which it was originally barred.
6. Its uses generalize and it is subject to further phonological reduction.
7. It eventually succumbs to competition from new forms and is eliminated.

This sequence is the reverse of that proposed by Hopper and Thompson (2008), who propose that the most colloquial versions of *wh*-clefts and ex-trapositives precede the more formal version and that their full forms are reformulations of the originals.

7. Conclusion

I have described the semantic and pragmatic meanings of TSs and shown how they are determined by the sentence’s lexis and syntax. I have also described several rhetorical uses to which the form may be put, and argued that these uses are due to the simultaneous presupposing and focusing of the TS clause. I have shown that TS variants represent a range of fixity and flexibility, in which some of the variants may be produced analytically or holistically, while others may be produced only holistically. I have shown that the register distribution of TSs as a type is skewed toward the informal end of the register spectrum, that is, to registers typically produced under

time pressure, which is where the less flexible variants most frequently occur. I have proposed a set of conditions for the creation of discourse management expressions like TSs and identified some of the linguistic elements from which they may be constructed and a diachronic trajectory for their origination, development, and extinction.

Colorado State University, USA

Notes

Corresponce address: Gerald.Delahunty@colostate.edu

1. I am grateful to Andreea Calude, Inga Vassilieva, Karen Vogel, the participants in the Perpignan conference on "Fixed Phrases in English," and two reviewers for the *Yearbook* for their helpful comments on earlier drafts of this paper. Any errors that remain are mine alone.
2. Example (6) occurs in the text of an interview regarding the education of girls in Benin, the relevant parts of which are:

Q: You grew up in Benin, ... Why is it so hard for girls to go to school?

A: *First, ... Secondly, ... and the third thing* is that ... So it's those three major things.

The thing is also money. Because school is not free. And when you're poor, you think about how much it's going to cost you, year after year, to put your child in school. And another reason also is the abuse of the girls in school

Because the complement of the highlighted sentence is a NP rather than a clause, it does not represent a propositional level discourse entity, as TS clauses do, so the discourse roles of the two complement types must necessarily differ. This sentence also comes after a series of numbered sub-topics, indicated by "First," "Secondly," and "the third thing," which are summed up at the end of Q's first full paragraph as "those three major things." "The thing" of the highlighted sentence may simply have been triggered by the several occurrences of "thing" in the immediately prior text, or it may contrast with "major things," and indicate a less important topic, or it may simply introduce an afterthought. Though this sentence seems to function as a marker of a shift in subtopic, as TSs often do, it does not reject and replace or revise an assumption derived from the prior context as TSs characteristically do (see Section 3 below), and so must be distinguished from them.

Example (7) also comes an interview.

Q: You've been deeply involved in bringing the Internet to India. How did this come about?

A: A key thing in my life has been not just computer communication, but computer communication for developing countries...

We were driven by the unarticulated needs of the developing market...

Another issue was language. Most e-mail around the world is in English. India has about 500 dialects and 17 scripts. What is great about technology – the fact that it is making our lives better is only part of it – is that it's giving us the hope of changing our lives. *The thing is to not be on the receiving end of fate, but to have your hand on the wheel – I think that changes you.* People needed something where they could post materials in their own languages.

To-infinitives and *that*-clauses are alike in that each represents a proposition, the latter more completely than the former. However, they differ semantically. *That*-clauses, unless they include a modal expression to the contrary, are factual, whereas *to*-infinitives denote “potential” (Quirk et al. 1985: 836), or “potentiality” or “projection into the future” (Huddleston and Pullum 2002: 1241). These elements of the interpretation of *to*-infinitives underlie the deontic interpretation so clearly present in (7), but lacking in the TSs with which this paper is concerned. Additionally, (7) does not reject and replace or revise an assumption licensed by the prior context.

Nonetheless, in spite of the difficulty of constructing a coherent interpretation of this paragraph, this sentence effects a shift in the trajectory of the topic, just as TSs and similar sentences with NP (and perhaps other) complements do, perhaps because of “The thing is ...” Clearly we are still at the initial stages of developing an understanding of the uses of *thing*.

3. This Spanish example shows that these sentence types occur in languages other than English, though the translation equivalents are not always discourse equivalent (Delahunty and Velazquez-Castillo 2002).
4. A reviewer for the *Yearbook* asked why I used COCA rather than BNC. COCA is a large (450 million words) balanced corpus consisting of the subcorpora I describe above (in this and other respects it is similar to the LSWE corpus used by Biber and his associates in the preparation of their *Longman Grammar of Spoken and Written English* [see pp. 29–35]), and as the spoken/written contrast seemed to be most significant in understanding the register distribution of TSs, COCA, with the limitations I describe in Section 2, being easily accessible and useable – and free – I opted for it.
5. A reviewer for the *Yearbook* asked that I amplify my assertion that “*thing* denotes merely COUNTABLE ENTITY.” The first synonym in the first definition of *thing* in the American Heritage Dictionary (4th edition) is “An entity, ...” (p. 1797), whose countability is clearly indicated by the indefinite article, cf. *stuff*. I gloss *thing* as COUNTABLE ENTITY to indicate that it is the most semantically general countable noun in the language, with the possible exception of *entity*. The reviewer also asked whether “my denotate [in the reviewer’s example from BNC] relate[s] to ‘things’ (plural, sic!) or something else? If it is the

latter, please explain." The reviewer's BNC example is, "it was strange then, all these things were happening, and the thing was, sometimes you didn't hear about them until much later." First, the clause in which *things* occurs is not a TS and without further context, I can only assume that *things* in this extract functions as a general noun (Halliday and Hasan 1976: ch. 6) referring to events – i.e., "things that were happening." Halliday and Hasan characterize general nouns as "a borderline case between a lexical item ... and a grammatical item ..." (p. 274). Second, because *thing* in the reviewer's example does occur in a TS, it functions very differently from *things*. It is part of a semantically minimal definite description which, by virtue of its function as the subject of a specificational sentence, denotes but does not refer. It is akin to the *it* of meteorological sentences, *there* of existential sentences, and other expressions which on some occasions are referential, and on others serve just some grammatical, metalinguistic, or discourse function. We might view *thing* as the nominal analog of "light verbs" whose role in collocations is addressed in Mel'čuk (1998: 31).

6. This is the only example I have found with a comma between *that* and the clause, suggesting that the matrix and clause even of full TSs may be intonationally rather than syntactically integrated.
7. Andreea Calude (p.c.) found one TS in the written and 60 TSs in the spoken subcorpora of the Wellington Corpus. Each subcorpus is one million words and the written instance is from fictional dialog. This cross-register distribution is similar to that in COCA, as I discuss below.
8. Regarding my claim that full TSs are fully consistent with the grammar, lexis, and semantics of English and therefore may be produced analytically rather than holistically, a reviewer for the *Yearbook* says, "it cannot be claimed that expressions which do not contain 'words or grammatical forms that are no longer current in the language' are not fixed or less fixed and less "frozen" than irregular expressions." This amounts to the claim that any expression may be fixed, even if it is entirely regular, grammatically, semantically, pragmatically, and discursively. I agree that this is possible. The reviewer also claims that "All TSs are constructions, regardless [of] their grammatical regularity." If s/he is using "constructions" in its broadest sense, as merely a grammatical form, then, of course, s/he is (tautologically) correct. If s/he is using the term as it is used in Construction Grammar, then s/he implies that TSs have linguistic and/or discourse properties that cannot be accounted for solely on the basis of their lexis, syntax, and semantics (cf. Mel'čuk's 1998: 30 quasi-phrasemes or quasi-idioms); however, I've shown that this is not the case for full TSs. If by "constructions," s/he merely intends that full TSs are fixed or formulaic to some degree, then the burden of proof is on those who would claim that an expression such as a full TS which superficially appears to be "unrestrictedly constructed" (cf. Mel'čuk 1998: 27) is fixed, and the discussion of the criteria for determining fixity in Section 4 shows just how

- difficult it is to support such a claim. The rare or non-occurrence of grammatically possible variants is just such evidence
9. *The thing is* without a comma or *that* seems to be intermediate between the full and more reduced and intonationally organized forms. It seems to be favored in moderately formal circumstances.
 10. *Newsweek* 10/18/10 p. 32, in an article on the puppet president of Chechnya, has *Trouble is, it's not always easy to see much difference*, perhaps reflecting the magazine's efforts to compete in a more colloquial market.
 11. Hudson (1998: 91–93) argues, following Hopper & Traugott (1993: 7), that this trajectory differs from that of grammaticalization, whereby words belonging to the major lexical or content classes become grammatical words, which become clitics, which become inflectional affixes.

References

- Aijmer, Karin. 2007. The interface between discourse and grammar: The fact is that. In Agnes Celle & Ruth Hart (eds.), *Connectives as discourse landmarks*, 31–46. Amsterdam: John Benjamins.
- Auer, Peter. 1992. Introduction: John Gumperz' approach to contextualization. In Peter Auer & Aldo di Luzio (eds.), *The contextualization of language*, 1–37. Amsterdam: John Benjamins.
- Bakhtin, Mikhail M. 1981[1953]. *The dialogic imagination* (ed. M. Holquist, tr. C. Emerson and M. Holquist). Austin: University of Texas Press.
- Bardovi-Harlig, Kathleen. 2009. Conventional expressions as a pragmalinguistic resource: Recognition and production of conventional expressions in L2 pragmatics. *Language Learning* 59. 755–795.
- Bergh, Gunnar. 2005. Mining the English language data on the Web: What can Google tell us? *ICAME Journal* 29. 25–46.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. Essex: Longman.
- Bond, Laurel. 2007. *Metaphors in context: Domains and lemmas as indicators of disciplinary differences in conceptualizations of "Reality."* Fort Collins, Colorado: Colorado State University MA thesis.
- Brenier, Jason M. & Laura A. Michaelis. 2005. Optimization via syntactic amalgam: Syntax-prosody mismatch and copula doubling. *Corpus Linguistics and Linguistic Theory* 1(1). 45–88.
- Brown, Penelope & Stephen Levinson. 1987. *Politeness: Some universals in language usage*. New York: Cambridge University Press.
- Carter, Ronald & Michael McCarthy. 2006. *Cambridge grammar of English*. New York: Cambridge University Press.
- Calude, Andreea & Gerald Delahunty. 2011. Inference in spoken English. *Pragmatics*. 21(3). 307–340.

- Coulmas, Florian. 1981. Introduction: Conversational routine. In Florian Coulmas (ed.), *Conversational routines: Explorations in standardized communication situations and prepatterned speech*, 1–17. The Hague: Mouton de Gruyter.
- Delahunty, Gerald. 2001. Discourse functions of inferential sentences. *Linguistics* 39. 517–545.
- Delahunty, Gerald. 2006. A relevance-theoretic analysis of *Not that* sentences. *Pragmatics* 16(2/3). 213–245.
- Delahunty, Gerald. 2007. A relevance-theoretic analysis of a lexico-syntactic pragmatic marker: *The thing is that S*. Paper presented at Hawaii International Conference on Arts and Humanities, 8 January.
- Delahunty, Gerald. 2008. What *Thing* sentences can tell us about register and mode. Paper presented at First North American Workshop on Pragmatics. Glendon College, York University, Toronto, Canada, 3–5 October.
- Delahunty, Gerald. In prep. An analysis of *The thing is that* sentences.
- Delahunty, Gerald & Maura Velazquez-Castillo. 2002. *The X is that S*: A lexicogrammatical device for local discourse management. In James F. Lee, Kimberly L. Geeslin & J. Clancy Clements (eds.), *Structure, meaning, and acquisition in Spanish*, 46–64. Somerville, MA: Cascadilla Press.
- Dor, Daniel. 2005. Toward a semantic account of *that*-deletion in English. *Linguistics* 43(2). 345–382.
- Edmonds, Amanda. 2010a. *On the representation of conventional expressions in L1-English and L2-French*. Bloomington, Indiana: Indiana University dissertation.
- Edmonds, Amanda. 2010b. Formulaic language and its many terms, definitions, and identification criteria. Paper presented at the International Conference on Fixed Phrases in English, Perpignan, France, 22–23 October.
- Finnell, A. 1996. *Discourse markers and topic change: A case study in historical pragmatics*. Cambridge, UK: University of Cambridge dissertation.
- Francis, Gill. 1994. Labelling discourse: An aspect of nominal-group lexical cohesion. In Malcolm Coulthard (ed.), *Advances in written text analysis*, 83–101. London: Routledge.
- Halliday, M.A.K. 1987. Spoken and written modes of meaning. In Rosalind Horowitz and S. Jay Samuels (eds.), *Comprehending oral and written language*, 55–82. San Diego: Academic Press.
- Halliday, M.A.K. & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Heller, Monica. 2001. Discourse and interaction. In Deborah Schiffrin, Deborah Tannen & Heidi E. Hamilton (eds.), *The handbook of discourse analysis*, 250–264. Malden, MA: Blackwell.
- Higgins, Francis R. 1976. *The pseudo-cleft construction in English*. Bloomington, Indiana: Indiana University Linguistics Club.
- Hopper, Paul & Sandra A. Thompson. 2008. Projectability and clause combining in interaction. In Ritva Laury (ed.), *Crosslinguistic studies of clause combining: The multifunctionality of conjunctions*, 99–123. Amsterdam: John Benjamins.

- Hopper, Paul & Elizabeth C. Traugott. 1993. *Grammaticalization*. Cambridge, UK: Cambridge University Press.
- Huddleston, Rodney & Geoffrey Pullum. 2002. *The Cambridge grammar of the English language*. New York: Cambridge University Press.
- Hudson, Jean. 1998. *Perspectives on fixedness: Applied and theoretical*. Lund: Lund University Press.
- Jespersen, Otto. 1976 [1926]. Living grammar. In *The philosophy of grammar*, 17–29. London: George Allen and Unwin. Reprinted in Diane D. Bornstein (ed.), *Readings in the theory of grammar*. 82–93. Cambridge, MA: Winthrop Publishers.
- Johnstone, Barbara. 2008. *Discourse analysis*, 2nd edn. Malden, MA: Blackwell.
- Kuiper, Koenraad, Marie-Elaine van Egmond, Gerard Kempen & Simone Sprenger. 2007. Slipping on superlemmas: Multi-word lexical items in speech production. *The Mental Lexicon* 2. 313–357.
- Lambrecht, Knud. 1994. *Information structure and sentence form*. Cambridge, UK: Cambridge University Press.
- Massam, Diane. 1999. *Thing is* constructions: The thing is, is what's the right analysis? *English Language and Linguistics* 3(2). 335–352.
- McConvell, Patrick. 1988. To be or double be? Current changes in the English copula. *Australian Journal of Linguistics* 8. 287–305.
- Mel'čuk, Igor. 1998. Collocation and lexical function. In Anthony P. Cowie (ed.), *Phraseology: Theory, analysis, and applications*, 23–53. Oxford: Clarendon Press.
- Moon, Rosamund. 1998. Frequencies and forms of phrasal lexemes in English. In Anthony P. Cowie (ed.), *Phraseology: Theory, analysis, and applications*, 79–100. Oxford: Clarendon Press.
- Pusch, Claus. 2006. Marqueurs discursifs et subordination syntaxique: La construction inférentielle en français et dans d'autres langues romanes. In M. Drescher & B. Frank-Job (eds.), *Les marqueurs discursifs dans les langues Romanes: Approches théoriques et méthodologiques*, 173–188. Frankfurt: Peter Lang.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. Essex, UK: Longman.
- Rankin, Ian. 2003. *Resurrection men*. Boston: Little, Brown, and Company.
- Rankin, Ian. 2005. *Fleshmarket alley*. New York: Time Warner.
- Schmid, Hans-Jörg. 2000. *English abstract nouns as conceptual shells: From corpus to cognition*. Berlin: Mouton de Gruyter.
- Sirr, Peter. 2009. The thing is. In *The thing is*. Loughcrew, Ireland: The Gallery Press.
- Tannen, Deborah. 1986. Introducing constructed dialogue in Greek and American conversational narratives. In Florian Coulmas (ed.), *Direct and indirect speech*, 311–32. New York: Mouton de Gruyter.
- Terasaki, Alene K. 2004 [1976]. Pre-announcement sequences in conversation. In Gene H. Lerner (ed.), *Conversation analysis: Studies from the first generation*, 171–224. Amsterdam: John Benjamins.

- Tuggy, David. 1996. The thing is is that people talk that way. The question is is Why? In Eugene H. Casad (ed.), *Cognitive linguistics in the redwoods: The expansion of a new paradigm in linguistics*, 713–752. Berlin: Mouton de Gruyter.
- Wilson, Deirdre & Dan Sperber. 2004. Relevance theory. In Laurence R. Horn and Gregory Ward (eds.), *The handbook of pragmatics*, 607–632. Malden, MA: Blackwell.
- Wray, Alison. 2002. *Formulaic language and the lexicon*. New York: Cambridge University Press.

