Transitivity Schemas of English EAT and DRINK in the BNC

John Newman and Sally Rice

1. Introduction

In a chapter called "The floating nature of transitiveness" in his multi-volume *Grammar of Late Modern English*, Hendrik Poutsma (1926: 54) wrote:

> Almost all verbs are used both transitively and intransitively.
> a) Sometimes the two applications appear to be equally natural, so that it would be difficult, or indeed impossible, to tell which is the original.
> b) Sometimes one application is clearly felt to be a modification of the other. It is especially this transition which is of particular interest to the student of English.

Transitivity, as a lexico-syntactic phenomenon, has long attracted attention in linguistics as an object of inquiry and as the subject of countless articles, monographs, and dissertations. The present study takes a very particular spin on the topic by examining and qualifying not the essence of transitivity, but merely the use of a verb as transitive or intransitive in particular contexts.[1] Moreover, we scrutinize only two verbs, the relatively basic English verbs EAT and DRINK.[2] However, we observe and quantify the transitivity of these two verbs across thousands of spoken and written examples from the British National Corpus (henceforth sBNC or wBNC to indicate the two sub-corpora, respectively, or BNC for the database as a whole) in the hopes of answering when and why these two highly volatile verbs enter into diathesis alternations—that is, alternations in the syntactic expression of arguments—in the first place.

The value of relying on the BNC as a source of diathesis alternation data can be appreciated in the (a) and (b) sentences in (1) and (2):

(1) a. *Well that put you in your place if you ate too many potatoes.* (sBNC)
b. *If I don't smoke, I eat.* (sBNC)

(2) a. *You just drank all my milk!* (sBNC)
b. *If you bet on horses or drink then it cost you money.* (sBNC)

We find that even such simple sentences as these are far more revealing as illustrative examples of the phenomenon under study than the examples found in the typical pedagogical grammar of English or treatise on theoretical syntax. One may note, for example, the presence of *too many* and *all* in the object phrases of the transitive (a) examples (indicative of a larger trend we established in our database); a preference for *I* and *you* as subjects in the spoken corpus; and an habitual use of the simple present in the intransitive (b) examples. Moreover, we feel that sentences like these can offer more insight into the polyvalency of verbs than the highly artificial and contrived examples in much of the literature with their requisite third person subjects, specific direct objects, simple present tense forms, and little in the way of adverbial modification.

The full range of diathesis alternations observed for EAT and DRINK in the BNC reveals that these verbs do indeed behave differently in both their argument structure and in the interpretation of their subject and object depending on genre, register, or modality (by which we mean spoken vs. written corpus). More significantly, we found that the overt valency of these verbs is strongly tied to their particular tense/aspect/mode (TAM) marking; the person, number, and specificity of their subject; as well as semantic properties of their propositional and extra-propositional collocates. In short, there is nothing binary nor straightforward about the so-called transitivity alternation. This paper summarizes the findings of our large-scale corpus inquiry on the grammatical patterning of EAT and DRINK. It also re-introduces a number of concerns about the structure and content of lexical entries (either theoretical or descriptive) as well as the 'floating' nature of transitivity itself.

The choice of EAT and DRINK as the focus of our study is not arbitrary. These items constitute a closely related pair of verbs within the same semantic domain,[3] comparable in their degree of (in)formality of usage, with each displaying the syntactic alternation of interest to us. In their uses with objects, they could be regarded as quintessential transitive verbs. In their uses with and without objects, they are the verbs of transitivity diathesis *par excellence*. In so far as they refer to bodily actions and everyday physiological experiences common to all humans, they could be called 'basic' verbs. As such, they are natural candidates for sources of figurative and metaphorical extension and idiomatic usage. They are obvious objects of interest and research for linguists with a cognitive linguistic orientation. Comparable research undertaken on other basic verbs from a cognitive linguistic viewpoint includes: sense-perception verbs (Sweetser 1990: 32–48); COME and GO (Radden 1996; Shen 1996;

Lichtenberk 1991); STAND and LIE (Serra Borneto 1996); SIT/STAND/LIE (Newman 2002; Newman and Rice 2004); SEE (Alm-Arvius 1993); GIVE and TAKE (Newman 1996, 1998); TAKE (Norvig and Lakoff 1987); HAVE (Wierzbicka 1988); and EAT and DRINK (Wierzbicka 1988; Newman 1997); and miscellaneous verbs referring to bodily acts (Pauwels and Simon-Vandenbergen 1995). For an overview of this and similar research see Newman (2004).

After reviewing some relevant proposals concerning transitivity in Section 2, we explain in Section 3 the corpus-based methodology that we have adopted for the purpose of this study. The distinction between spoken versus written modalities pervades our discussion and we consider some larger findings in terms of modality differences in Section 4. We examine in Sections 5 and 6 object and subject phrases in greater detail and discuss—where appropriate—the relevance of our results to claims about transitivity.

2. Models of transitivity alternations

In the approach adopted here, we understand 'transitive' in a conservative manner. 'Transitive' designates a construction in which a verb is used with a direct object, whereas 'intransitive' refers to a construction in which a verb is used without one. While neither 'construction' nor 'direct object' is unproblematic as a theoretical term, these labels are nevertheless useful in helping us to delineate the intended sense of transitivity. The linguistics literature offers quite a range of interpretations. Huddleston (1988: 59-60) happens to illustrate the traditional view of transitivity (and omitted object constructions) with an EAT example, repeated here as (3), where S = subject, P = predicate, and O^d = direct object.

(3)	a.	*She ate.*	Intransitive	S P
	b.	*She ate an apple.*	Monotransitive	S P O^d

The sentence in (3a) is deemed intransitive and not further distinguished from what we might recognize as a traditional intransitive like *She died*. Huddleston (1988: 60) describes the propositional relationship in (3a) in the following way: "…the participant role of the subject-referent remains constant and the intransitive clause simply leaves unexpressed the second participant. *She ate* entails that she ate something but doesn't specify what." So-called monotransitives like (3b) receive little in the way of

further analysis. Rather, it is the intransitive alternate in (3a) which attracts all the attention. We believe that each of the constructions represented in (3) is worthy of study in its own right and neither is derivative of the other.

Huddleston and Pullum (2002: 303-305) refine Huddleston's notion of intransitivity by offering a sub-categorization of types of 'unexpressed objects' of intransitive verbs. EAT and DRINK participate in two such patterns of omissibility: 'specific category indefinites' and 'normal category indefinites'. The former refers to the possibility of understanding the intransitive uses of EAT and DRINK specifically as 'eat a meal' and 'drink alcoholic drink' respectively; the latter refers to the use of intransitive EAT and DRINK when the unexpressed object is interpreted as the 'indefinite, typical, unexceptional' exemplar ('food' in the case of EAT, 'water' or 'beverage' presumably, in the case of DRINK).

The traditional view of an intransitive vs. (mono)transitive distinction, as enunciated in Huddleston (1988) and Huddleston and Pullum (2002), is by no means compelling. One could just as well argue that the intransitive use in (3a) really involves one participant (the agent phrase) and describes an activity of that participant, similar to the way in which the intransitive verb *run* in English describes an activity of a runner. Other associated entities can be a necessary part of a larger semantic frame of intransitive verbs (legs in the case of *run*, food in the case of *eat*), but this does not require us to say that they are second participants which are simply unexpressed. In our discussion below, we investigate properties of intransitive and transitive uses of verbs separately, without any assumption that the intransitive use is reducible to the transitive use with the direct object unexpressed. We regard intransitive and transitive uses of EAT and DRINK as separate constructions, or *schemas*, with a host of quite different properties. These schemas are associated with preferred kinds of subjects and objects in terms of both grammatical and lexical content and with preferred co-occurrence patterns of subject, object, and TAM marking.

A more provocative view of transitivity can be found in Van Valin and LaPolla (1997: 115). They speak of the English predicate as having either one or two arguments in its logical structure, similar to Huddleston's distinction between intransitive and monotransitive uses of EAT. Their representation of the logical form of EAT expresses the alternatives through the parenthesized (y) embedded in the argument structure.

(4) **do'** (x, [**eat'** (x, (y))]
x=CONSUMER, y=CONSUMED

Van Valin and LaPolla distinguish the semantic roles as found in logical structure (agent, patient, etc.) and what they call 'macroroles' (actor, undergoer). Applying a notion of transitivity at the level of macroroles ('M-transitivity'), they draw a three-way distinction between *atransitive*, *intransitive*, and *transitive* verb types, as shown in Table 1.

Table 1. An illustration of M-transitivity, based on Van Valin and LaPolla (1997; 99, 147, 150) (eat_1 is considered an activity verb, while eat_2 is an active accomplishment verb)

verb	example sentence	semantic valence	macrorole number	M-transitivity
rain	*It rained.*	0	0	Atransitive
eat_1	*He ate.*	1 or 2	1	Intransitive
	He ate spaghetti for ten minutes.			
eat_2	*He ate a plate of spaghetti in ten minutes.*	2	2	Transitive

There is a partial overlap with the traditional notion of transitivity in so far as *He ate* is intransitive in both systems and *He ate the plate of spaghetti in ten minutes* is transitive in both systems. A non-referential, 'inherent' argument, as found in *He ate spaghetti for ten minutes*, however, does not have an undergoer macrorole assigned to it. Verbs with such inherent arguments are 'intransitive' in terms of M-transitivity in Van Valin and LaPolla (1997: 147-154). A search in the sBNC shows, incidentally, that the use of non-referential objects with *ate* is relatively rare: of 155 instances of the verb form *ate* in the BNC, in only three cases do we find non-referential objects: *And it turned out that there was a big goblin that lived on this island and he just ate fairies; Mind you, I ate conga; They ate boiled eggs for breakfast.*

In distinguishing referential and non-referential kinds of arguments—a distinction with important morphosyntactic ramifications in some languages—Van Valin and LaPolla achieve a certain refinement of the concept of transitivity, though it is at odds with the traditional account. Not only are the M-transitive and M-intransitive classes not identical with their traditional counterparts, the M-intransitive uses of EAT are not to be understood as simply reduced versions of the M-transitive uses (a view we also endorse). Van Valin and LaPolla (1997:112) explicitly remark that

"…*eat* is not inherently telic, unlike *kill* and *break*; hence it must be analyzed as an activity verb, with an active accomplishment use". For them, the 'activity verb' use (*He ate*, *He ate spaghetti for ten minutes*) is the 'basic' meaning of EAT. The examples used by Van Valin and LaPolla to illustrate EAT used as an activity verb and as an activity accomplishment verb are, of course, constructed examples. Again, a search of the sBNC can offer insights into the naturalness, or lack thereof, in having a *for*-phrase in their examples containing syntactic objects in Table 1. Again, of the 155 instances of *ate* in the sBNC, there is no example of any usage which conforms to the constructional pattern [Subject-NP *ate* Object-NP *for* Time-Phrase], whereas the construction [Subject-NP *ate* Object-NP *in* Time-Phrase] is attested in a couple of examples, e.g., *Six swallows ate three hundred flies in five hours*.

An influential and far-reaching re-conceptualization of the notion of transitivity is that found in Hopper and Thompson (1980). They identify 10 parameters which distinguish high and low transitivity of clauses, as summarized in Table 2.

Table 2. Scalar transitivity according to Hopper and Thompson (1980: 252)

	HIGH	LOW
A. PARTICIPANTS	2	1
B. KINESIS	action	non-action
C. ASPECT	telic	atelic
D. PUNCTUALITY	punctual	non-punctual
E. VOLITIONALITY	volitional	non-volitional
F. AFFIRMATION	affirmative	negative
G. MODE	realis	irrealis
H. AGENCY	A high in potency	A low in potency
I. AFFECTEDNESS OF O	O highly affected	O not affected
J. INDIVIDUATION OF O	O highly individuated	O not individuated

The high values for these parameters are claimed to co-vary with one another within a language and cross-linguistically; similarly for the low values. While Hopper and Thompson are not directly concerned with the kind of transitive/intransitive alternation found with EAT and DRINK, their account of transitivity—a clause-level phenomenon for them, rather than simply a verbal one—posits a scale of transitivity allowing for varying degrees to which each of the 10 parameters could be said to be either 'high' or 'low'. Hopper and Thompson's work suggests that there are additional distinctions that might profitably be drawn amongst the syntactically

transitive EAT/DRINK uses (in terms of kinesis, aspect, etc.), just as there are additional distinctions that can be made amongst the syntactically intransitive EAT/DRINK uses. Presence or absence of a syntactic object, in other words, is not the only factor of relevance in considering the Transitivity of a clause (written here with a capital T to indicate the Hopper and Thompson sense of transitivity); it is merely one of 10 Transitivity parameters that might be examined.[4]

Recently, Thompson and Hopper (2001) have revisited Transitivity through an exploration of the syntactic patterning found in spontaneous conversation. They arrive at a number of conclusions with relevance to the present study. One result from their corpus-based research is that Transitivity, understood as a kind of additive phenomenon with respect to the number and value of the parameters in Table 2, is very low in spontaneous conversation. This result immediately suggested a line of inquiry with respect to EAT and DRINK and consequently we applied selected Hopper and Thompson (1980) Transitivity parameters to our search returns from the BNC. Thompson and Hopper (2001: 43) acknowledge that, even within a language, there will be variation amongst verbs with respect to their parameters. Thus, the specific behaviour of EAT and DRINK needs to be determined in its own right for each of these verbs, even if we accept their claim that, overall, Transitivity has been found to be low in spontaneous conversation. Thompson and Hopper (2001) is instructive, too, in its usage-based approach to investigating Transitivity. Their focus was spontaneous conversation and their methodology involves a close examination of real examples taken from interactional communication. In other words, they adopt a corpus-linguistic approach and arrive at usage-based results which hold true for a particular genre. In so doing, they distinguish themselves from traditional approaches to grammar which are neither corpus-based nor particularly sensitive to genre or modality differences.[5]

The recent and prestigious *Cambridge Grammar of the English Language* by Huddleston and Pullum (2002) already cited above is typical of this tradition, though perhaps it is more explicit about its assumptions than most other grammars. The authors aim for a grammatical description which is "neutral between spoken and written English" (ibid.: 11). They justify this approach with the claim that "[s]harp divergences between the syntax of speech and the syntax of writing, as opposed to differences that exist between styles within either the spoken or the written language, are rare to the point of non-existence" (ibid.: 13). We believe there is sufficient evidence to justify modality-specific grammatical descriptions and, in the

approach adopted here, the contrast in the morphosyntactic behavior of these two key verbs in spoken vs. written modality is a pervasive and crucial feature of our analyses. Our corpus-based approach and the findings which we turn to next ensure a naturalness in our examples, as opposed to the constructed examples of a grammar such as Huddleston and Pullum (2002)[6].

3. Methodology

The British National Corpus World Edition (BNC) was used as the basis for all searches discussed below. Initially, the entire spoken sub-corpus of the BNC (10 million words) was searched for the word forms *eat, eats, eating, ate, eaten* and *drink, drinks, drinking, drank, drunk*, without recourse to tags. All these results were saved (2,623 hits for EAT, 934 hits for DRINK). The written sub-corpus of the BNC (90 million words) was searched using the same keyword list, but owing to the vast size of the corpus, only a random sampling of 2,000 hits for each of EAT and DRINK was saved. These search returns—the exhaustive sBNC search and what we hope is a representative wBNC search—form the basis of the analysis presented here.

The 7,557 (= 2,623 + 934 + 2,000 + 2,000) examples of EAT and DRINK downloaded from the BNC and used in our analyses were exported to a FileMaker Pro™ database where each return was individually examined and coded as its own record. To begin with, each record was checked for whether or not it represented a form of the relevant verb. Thus, the adjective *drunk* 'intoxicated' was excluded from further analysis, as were the nominal uses of *drink*, *drinks*, *eats*, and *EAT* (Employment Appeal Tribunal). We excluded *–ing* forms in compounds such as *eating habits*, *eating disorders*, *eating places*, *drinking fountains,* etc. Each legitimate verb usage was coded for its source corpus (sBNC or wBNC), the part of speech and inflectional class of the key word, and whether the usage had an overt direct object (transitive) or not (intransitive). Moreover, both the subject NP and the direct object, if present, were identified in separate fields. We were thus able to quantify the exact number of usages, for example, of *had eaten* with a first person plural subject and an omitted object or all instances of *drinks* with an unspecified third person subject and overt object.

A small proportion of the examples were deemed uninterpretable (and these were linked exclusively to the sBNC). We downloaded the keyword

in our searches, e.g., *eats*, with a limited amount of left and right context (40 characters on each side of the keyword for the sBNC) although in some cases the context may not have been large enough to retrieve the information we wished to identify. However, it is a fact about spontaneous conversation that there will be interruptions, false starts, incomplete utterances, and so forth, and even looking up the full context of usage may not yield information about the subject, object, etc. We adopted a conservative approach to the interpretation of the returns from the sBNC, coding uses as 'uninterpretable' unless we were very confident about the meaning. A less conservative approach might have classified a number of these instances as indeed interpretable. It is unlikely that the subsequent subclassification of these unclear instances (e.g., determining the relative proportions of 1SG, 2SG/PL, 3SG subjects) would be seriously affected by a different stance with respect to their inclusion. The examples in (5) illustrate some of our 'uninterpretable' instances, these being cases where we were unable to confidently identify a subject of EAT or DRINK.

(5) Examples of an unidentifiable or uninterpretable subject in the coding (key verb underlined)

a. *Potatoes and er Yeah. Sussed it! bread and eat Yeah. things like that.*
b. *Yeah, but you can also get it with salt mm Can't eat popcorn then.*
c. *No I don't think you'll have both Why not? eat it in the evening?*
d. *Yeah. I thought it'd have something eat it. Your joking! Well erm hello!*
e. *Baa, baa black. Come on stop it, supposed to be eating. Baa.*
f. *Well he went to someone's house and all the eats hem so he ate it.*
g. *Oh dear. And did she take her cup out? Yeah drink about tea.*
h. *Er this fellow come up fair blue devil go drinking, er he hadn't a, he hadn't.*

While the identification of an object phrase presented few difficulties, identifying and coding the subject phrase posed a number of problems. We distinguished 'specified' and 'unspecified' subjects, consistent with a contemporary linguistic approach. Even so, it was necessary to make decisions relating to just how narrow or broad our categorizations were going to be. We took 'specified subject' to be an overt phrase functioning as the subject of a finite clause containing EAT or DRINK in the main verbal complex or as the phrase functioning as the understood subject of a nearby

participial or infinitival phrase. Specified subjects were then sub-categorized for number and person. The examples in (6), taken from our database, illustrate a variety of specified subject phrases of EAT and DRINK.

(6) Examples of a specified subject in the coding (subject phrase underlined)

a. *She felt she might never eat again.*
b. *He managed to eat most of the cream.*
c. *Can I mummy? No You're not big enough to drink wine...*
d. *…when you first start to drink spirits you feel Oh dear, Yeah.*
e. *They used to cut it up and pretend to eat it.*
f. *…made him promise never to eat again.*
g. *…watching other monkeys trying to eat these insects.*
h. *…but then I hate drinking milk anyway.*
i. *You'll soon get fed up eating it.*
j. *Next to me a girl eating a box of chocolates…*
k. *He was sitting there eating…*
l. *…the poor, who live, cook, eat, and sleep…*

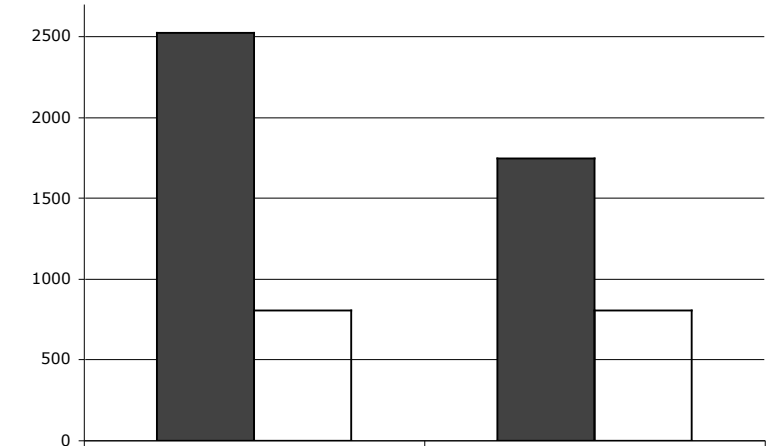
'Unspecified subject', on the other hand, was taken to refer to an unexpressed, but understood, generic agent associated with an infinitival or *–ing* form. The examples in (7) illustrate such cases with *eat* and *eating*. Sometimes, it is possible in the unspecified subject constructions to relate the eating or drinking to a person referred to elsewhere in the sentence or larger context. For example, in (7b), it is obviously Annie who is the intended eater of the currants. Although it is possible in many such cases to identify a specific agent, we took the construction itself to have an unspecified subject, despite the inferences the larger context allows.

(7) Examples of an unspecified subject in the coding (the relevant verb form is underlined)

a. *I haven't had a thing to eat for hours.*
b. *Carolyn gave Annie a saucer of currants to eat.*
c. *…in case we don't encounter a suitable place to eat.*
d. *To eat chalk is as foolish as to try to write on a blackboard with cheese!*
e. *But I would always ask, Is it safe to drink the water out the taps?*
f. *The best way to protect the pig is to eat it.*
g. *Eating apples is good.*
h. *It claimed Elton John was hooked on eating food and spitting it out.*
i. *But, sad to say, talking and drinking got the better of him.*

4. Spoken vs. written results

As a point of departure, let us look at some of the most striking and macro-level results from our BNC searches. Figure 1 compares the number of EAT and DRINK verbs in our database. In this table, as throughout, the written corpus is the *sampled* wBNC, obtained as described in the preceding section and one should be wary of making direct comparisons between the raw numbers in our spoken and written samples. Figure 1 includes specified and unspecified verb forms, as well as the verb forms which had 'uninterpretable' subjects, but excludes non-verb forms of EAT and DRINK.



	SPOKEN	WRITTEN
EAT	2526	1748
DRINK	807	804

Figure 1. The incidence of EAT and DRINK verbs by sub-corpus of the BNC

As can be readily seen in Figure 1, there is a clear preponderance of EAT forms over DRINK in both the spoken and written corpora. The higher frequency of EAT is only one way in which it is more salient than DRINK. EAT also has special status vis-à-vis DRINK in terms of the relative order one tends to use in describing the combination of the two types of consumption: *eating and drinking*, rather than *drinking and eating*. To corroborate this intuition about sequential ordering of EAT and DRINK words, we conducted a series of searches on the conjunction of the inflected forms of EAT and DRINK in the whole BNC (*eat and drink*, *drink and eat*, *eats and drinks*, etc.). The results of these searches, shown in

Table 3, confirm this preference in the relative ordering of EAT and DRINK in conjoined phrases. With the exception of the second row, the results are all highly significant as determined by two-tailed binomial tests ($p<0.0001$, $p=0.625$, $p<0.0001$, $p<0.0001$). The results are reminiscent of what we found elsewhere with the verb set SIT, STAND, and LIE, where corpus research shows a relative frequency of SIT > STAND > LIE (cf. Newman and Rice 2001), matching a preference for the same order in phrasal combinations. Higher frequency and priority in sequential ordering are both potential indicators of experiential salience: when we eat and drink, the drinking is an accompaniment to the eating, rather than the other way around. The preferred order of EAT and DRINK words could be seen as an instance of the more general Food and Drink Hierarchy (Fish > Meat > Drink, etc.) proposed by Cooper and Ross (1975), though Cooper and Ross based their hierarchy on nouns rather than verbs.

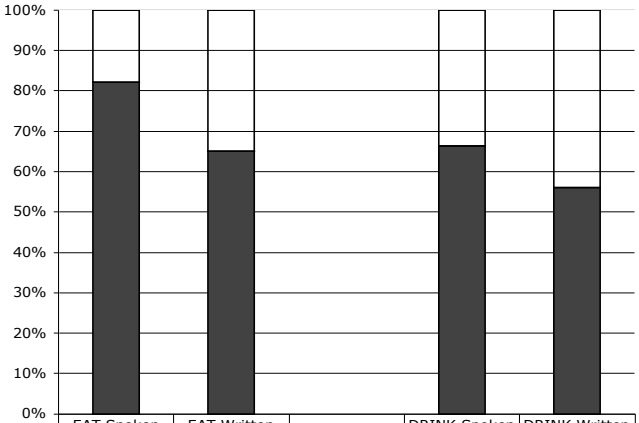
Table 3. Frequencies of conjoined EAT and DRINK in the whole BNC

eat and drink	66	*drink and eat*	2
eats and drinks	3	*drinks and eats*	1
ate and drank	25	*drank and ate*	2
eating and drinking	70	*drinking and eating*	8

Though sceptical based on our previous corpus research (cf. Newman and Rice 2001, Rice and Newman 2004, and Newman and Rice 2004), we were initially mindful of Huddleston and Pullum's pronouncement, quoted above, that the syntax of spoken and written language is virtually the same. We took this as a sort of null hypothesis as we began our investigation of the rather robust diathesis alternations affecting EAT and DRINK across modalities. Admittedly, we were more sympathetic to Thompson and Hopper's claim that Transitivity is very low in spontaneous conversation and expected the incidence of overt vs. omitted objects for these two verbs to vary greatly by corpus. As it happens, our results contravened both accounts.

As shown in Figure 2, not only were these verbs being used transitively, with full-blown objects, most of the time, the incidence of transitive usages was greatest in the spoken corpus for both verbs. In short, pace Huddleston and Pullum, the relative distribution between transitive and intransitive usages—whatever it is—is not consistent across modalities. Figure 2 presents lemmatized totals for these verbs across the two contrasting argument structures. Later, we will give totals for the relative distribution across inflectional forms (by both person and number of

subject and TAM marking on the verb), argument structures, and sub-corpus. While one might wish for a relevant measure of statisical significance in evaluating Figure 2, it is not at all clear what the appropriate measure would be. Recall that the "written" counts are based on samples from the wBNC (2,000 hits for each of the lemmas EAT and DRINK), whereas the "spoken" counts are based on total occurrences in the whole of the sBNC (10 million words). Familiar statistical tests such as chi square appear to be inappropriate in this case, given the discrepancy in the nature and size of the written and spoken results on which the comparison is based (cf. Kilgarriff 2001: 124).



	EAT Spoken	EAT Written		DRINK Spoken	DRINK Written
INTRANSITIVE	451	610		271	353
TRANSITIVE	2075	1138		536	451

Figure 2. Percentage of transitive and intransitive usages of EAT and DRINK by sub-corpus of the BNC

The findings summarized in Figure 2 also appear to contradict the results of Thompson and Hopper (2001) who claim that low Transitivity is a feature of spontaneous conversation. Some qualifications concerning this comparison are in order, however. Firstly, Transitivity is a composite of 10 parameters, only one of which relates to the presence of an object (parameter A, participants). The transitivity reported on in Table 5, on the other hand, is the traditional notion and concerns merely the presence or absence of an object. Secondly, one cannot equate Thompson and Hopper's

2-participant clauses with the presence of an object in the BNC cases. A transitive use of EAT or DRINK, for example, may occur without any expressed subject (unspecified subject constructions, imperatives, etc.), in which case the clause would count as a 1-participant clause. Thirdly, it should be remembered that Thompson and Hopper base their conclusions on spontaneous conversation only, whereas the sBNC includes a variety of genres (monologue, dialogue) and domains (educational, business, public, leisure). While these are real considerations, our results on transitivity would still appear to be at odds with Thompson and Hopper (2001).

The difference in the behaviors of EAT and DRINK is also noteworthy. There is proportionately more intransitive usage with DRINK than there is with EAT. The difference is arguably influenced by the existence of specialized meanings associated with the intransitive (the 'specific category indefinite' kind of interpretation à la Huddleston and Pullum 2002: 303-305 or Rice 1988). In the case of EAT the specific interpretation is 'meal', whereas with DRINK it is 'alcoholic beverage' (especially when consumed in an habitual and/or excessive manner). This use of intransitive DRINK is a very familiar one in casual conversation (some examples from sBNC are *All they do in that house is drink and smoke*; *Because her daddy drinks in there in the pub...*; *He bought a bottle of brandy at the first liquor store he found and he began to drink*), reflecting the prominence of alchol consumption as a topic of discourse. Comparing EAT and DRINK in this way is instructive for demonstrating the kind of variation that can exist between lexical items, even those which define and exhaust a class (cf. Levin 1993: 213-214). The variation becomes more pronounced in the next set of results.

Figure 2 above only summarizes the relative frequencies of transitive and intransitive usage, ignoring the variation that exists between modalities and between different subject choices. By contrast, Figures 3 and 4 show more detail of this variation in the relative frequencies of transitive and intransitive uses for the lexical forms (not the lemmas) *eat* and *drink*, respectively.[7] These are the forms which occur as finite verbs (*We eat dinner at 6.00pm*), as well as infinitival forms (*We like to eat dinner late*, *We may eat dinner late*, *There's too much to eat*, etc.) and imperative forms. While there is, overall, a higher percentage of transitive than intransitive usage for EAT and DRINK, these figures reveal the varying percentages evident in more specific inflections. Indeed, the intransitive usage is the dominant usage in some instances. In the case of *eat*, for example, the intransitive usage is greater with first person plural (1p) subjects in both spoken and written corpora. With *drink*, on the other hand,

transitive usage is greater with 1p subjects. In the case of third person plural (3p) subjects, the intransitive use of both *eat* and *drink* is also greater than the transitive use in the spoken corpus. It can be seen in Figures 3 and 4 that there is a preference for intransitive use of *eat* with (1st and 3rd person) plural subjects. Possibly, the experiential realities of eating limit the range of possible objects of *eat* with plural subjects (it is more natural for one person to eat a specific item of food than it is for a group of people to do so). Figures 3 and 4 are meant to give some sense of the variation that exists across different subjects, between the corpora, and between the two lexical items (see Appendices 1 and 2 for the frequencies and percentages underlying this figure). While there is an abundance of similar results that one could generate from our database, we content ourselves here with exemplifying the considerable variation that is lurking within Figure 2.

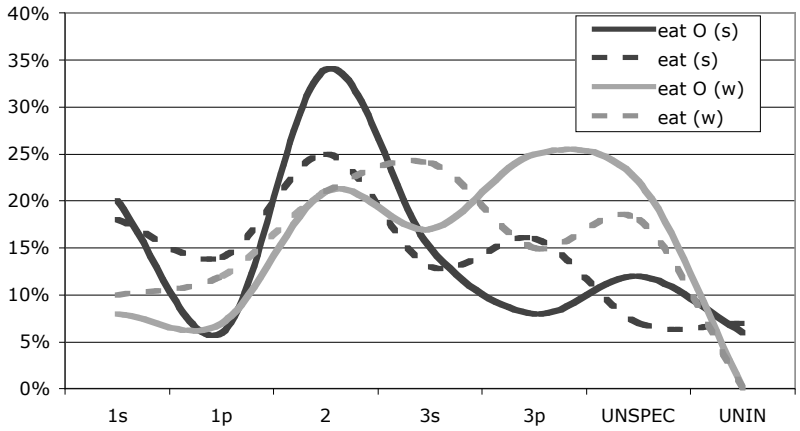


Figure 3. Percentage use of *eat* with (solid lines) and without (broken lines) an object in the sBNC and sampled wBNC

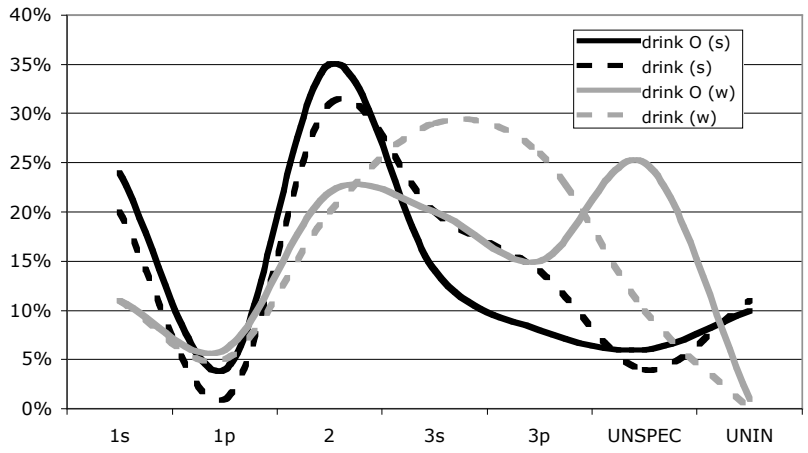


Figure 4. Percentage use of *drink* with (solid lines) and without (broken lines) an object in the sBNC and sampled wBNC

5. Objects

In this section we examine in more detail properties of the objects of EAT and DRINK. We will look at a number of properties which can be successfully researched and described in a quantifiable way, consistent with the orientation of the present study.

5.1 Individuation of O

We explored the degree of individuation of the object (parameter J for Hopper and Thompson 1980) evident in our data from sBNC and the sampled wBNC. Pronouns are high in individuation and since they are easily identified and searched in a corpus, they are an effective means of measuring individuation. Thompson and Hopper (2001: 36), for example, determine the frequency of pronominal objects in their corpus as a way of quantifying individuation. We followed a similar methodology, though we relied entirely on electronic corpus linguistic tools, as is necessary when working with a database of the size we were dealing with. We used Wordsmith™ to produce frequency lists of the individual words occurring

within the object phrases of EAT and DRINK. Our procedure did not differentiate between a single pronoun as object (as in *eat it*) and the relatively infrequent case of a pronoun occurring in a modifying phrase within the object (as in *eat some of it*). This indeterminancy about the syntactic status of the pronouns within the object field in our database is a small cost for the larger benefit of using automated frequency counts with a large database.

We present the results of the object listings in Tables 4 and 5 in terms of frequency-ranking, i.e., the rank occupied by an item in terms of its frequency of occurrence in the domain of the search (in this case, the object field). The highest rank is occupied by the most frequently occurring word (or set of words if the words have identical frequencies). In the case of the lemma EAT, both *it* and *them* appear in the top ten rankings in both the sBNC and the wBNC. However, in both cases, their ranking is higher in the sBNC than in the wBNC. In the case of EAT, which comfortably allows for singular or plural objects, we find higher rankings in the sBNC compared with the wBNC (rank 1 vs. rank 5 for *it*, rank 3 vs. rank 9 for *them*). An even more striking difference is found with the lemma DRINK where only *it*, and not *them*, occurs in the top 10 rankings. There is an experiential rationale for this: we drink liquids which are commonly referred to by mass nouns (hence, treated as singular). Furthermore, we tend to drink one type of beverage at a time. We find that *it* is, in fact, the single most frequently occurring word in the object position with DRINK in the sBNC.

Table 4. The 10 highest frequency-rankings of single words in the object phrases of EAT in the sBNC and wBNC (*it* and *them* in bold)

	EAT (sBNC)			EAT (wBNC sample)	
rank	object keywords	N	rank	object keywords	N
1	***it***	**375**	1	*the*	157
2	*the*	189	2	*of*	136
3	***them***	**155**	3	*a*	133
4	*all*	148	4	*and*	86
5	*that*	133	5	***it***	**82**
6	*of*	132	6	*what*	67
7	*a*	109	7	*food*	65
8	*what*	102	8	*much*	38
9	*your*	83	8	*something*	38
10	*something*	76	9	***them***	**31**
			10	*foods*	29
			10	*more*	29

Table 5. The 10 highest frequency-rankings of single words in the object phrases of DRINK in the sBNC and wBNC (*it* in bold)

	DRINK (sBNC)			DRINK (wBNC sample)	
rank	object keywords	N	rank	object keywords	N
1	***it***	**104**	1	*of*	79
2	*a*	57	2	*the*	58
3	*that*	51	3	*a*	57
4	*of*	47	4	*much*	51
5	*tea*	46	5	*coffee*	43
6	*coffee*	41	6	*tea*	39
7	*what*	31	7	*wine*	33
8	*the*	28	8	*water*	32
8	*your*	28	9	*too*	31
9	*much*	25	10	***it***	**27**
10	*lot*	23			
10	*milk*	23			

These results clearly indicate a higher individuation of objects in the spoken modality compared with the written, as well as a certain difference in this Transitivity parameter between EAT and DRINK. The higher individuation with EAT and DRINK in the spoken corpus would appear to contradict Thompson and Hopper's claims about Transitivity being

relatively low in spontaneous conversation. However, even for them, the individuation parameter did not show up as a high value amongst objects in 2-participant clauses. It was one of three parameters which they describe as being divided more or less evenly between high and low values in spontaneous conversation, the other two being volitionality and mode. Our results, however, do not support an even balance of individuation between spoken and written modalities; rather, we find the spoken modality favoring high individuation. The same qualification needs to be made here as above concerning the genres examined by Thompson and Hopper and those in the present study. They base their conclusions on spontaneous conversation only, whereas the sBNC includes a variety of genres (monologue, dialogue) and domains (educational, business, public, leisure).

5.2 Affectedness of O

We also examined the affectedness of O (parameter I for Hopper and Thompson) by considering the type and frequency of 'excessive' modifiers or quantifiers that appeared in the object phrase. To help us efficiently identify recurring patterns, we obtained trigrams, or 3-word clusters, from the object phrases. Trigrams typically include combinations of full lexical items and functional words, e.g., *a lot of*, *a cup of*, *cup of tea,* etc. As such, they reveal more of the affectedness of the object than do bigrams, which will include combinations of function words only, e.g., *in a, of the*. Trigrams within the objects were calculated separately for EAT and DRINK for each of the spoken and written corpora. Tables 6 and 7 summarize the results of this operation, showing the top 20 trigrams as listed by Wordsmith™. 'Top 20' refers to the first 20 trigrams (in order of decreasing frequency) which appear in the list of word clusters compiled by Wordsmith™. There may be additional trigrams (in the 21st, 22nd position, etc.) with identical frequencies as the 20th trigram in some of these lists but, for the sake of ease of comparison of results, these are not included in the tables. As happens with n-gram analyses, some word sequences will appear as separate n-grams when, in fact, they are overlapping sub-parts of a larger construction. So, for example, both *twice as much* and *as much as* will be treated as separate trigrams in the phrase *twice as much as*. Some overlapping of this sort is evident in these tables.

Table 6. Top 20 trigrams from object phrases of EAT in the sBNC and wBNC (trigrams with 'excessive' descriptors in bold)

	EAT (sBNC)		EAT (wBNC sample)	
rank	top trigrams	N	top trigrams	N
1	***a lot of***	**14**	***as much as***	**7**
2	***as much as***	**10**	***twice as much***	**4**
3	*a little bit*	5	*a dish of*	3
4	*kind of things*	5	***a lot of***	**3**
5	*one of these*	5	*a piece of*	3
6	*sort of thing*	4	*bread and cheese*	3
7	***a bit more***	**3**	*fruit and vegetables*	3
8	*a couple of*	3	***most of the***	**3**
9	***all of them***	**3**	*some of the*	3
10	*any more of*	3	*a bar of*	2
11	*as you like*	3	*a healthy diet*	2
12	***little bit more***	**3**	*a healthy well*	2
13	***more of that***	**3**	*bar of chocolate*	2
14	***most of the***	**3**	*bread and jam*	2
15	*of the things*	3	*fish and chips*	2
16	*one of them*	3	*foods rich in*	2
17	*one of those*	3	*good country food*	2
18	***quite a lot***	**3**	*healthy well-balanced*	2
19	*the pink bits*	3	*kind of food*	2
20	*three hundred flies*	3	***large amounts of***	**2**

The trigrams in these two tables reveal a propensity towards lexical items relating to an increased or excessive degree of consumption for both EAT and DRINK, as found in each of the corpora though more so in the sBNC. The relevant trigrams are shown in bold in these tables: *a lot of*, *all of them*, *most of the*, *more of the*, *loads and loads*, *endless cups of*, *large amounts of*, etc. We consider these results to be of some interest in that they draw attention to a pattern of usage of transitive EAT and DRINK which is rarely acknowledged, for example, in dictionaries. It is well-known that the intransitive usage of DRINK has associations of an habitual and excessive consumption of alcohol, a meaning regularly recognized in dictionaries. But the idea of excessiveness is also salient in the overt, expressed object phrases of transitive DRINK, as it is for EAT. Likewise, there are more excessive descriptors with DRINK in the sBNC than in the sampled wBNC. This finding should be qualified, however, since quantifier phrases containing 'excessive' notions, such as *a lot of* are well represented

in spoken corpora generally, especially in comparison to written (cf. Biber et al. 2000: 277-278).

Table 7. Top 20 trigrams from object phrases of DRINK in the sBNC and wBNC (trigrams with 'excessive' descriptors in bold)

	DRINK (sBNC)		DRINK (wBNC sample)	
rank	top trigrams	N	top trigrams	N
1	***a lot of***	**8**	*a cup of*	5
2	*a bottle of*	6	***a little too (much)***	**5**
3	*a cup of*	5	***little too much***	**5**
4	*cup of tea*	5	*a bottle of*	4
5	*your orange juice*	4	*cup of tea*	4
6	*a pint of*	3	*a mug of*	3
7	*bottle of wine*	3	***endless cups of***	**3**
8	*cup of coffee*	3	*or fruit juice*	3
9	***lot of beer***	**3**	*water or fruit*	3
10	***lot of it***	**3**	*a litre of*	2
11	*bottle of gin*	2	***a lot of***	**2**
12	*drop of milk*	2	*a pint of*	2
13	***gallons of it***	**2**	***as much as***	**2**
14	*little drop of*	2	*cups of coffee*	2
15	***loads and loads***	**2**	*cups of tea*	2
16	*red hot stuff*	2	*half a bottle*	2
17	***too much coffee***	**2**	*mineral water or*	2
18	*two or three*	2	*mug of tea*	2
19			*one of the*	2
20			*pints of lager*	5

5.3 Preferred objects of consumption

Since one of our goals was to better understand the nature of the overt objects occurring with EAT and DRINK, we identified the most common kinds of nouns referring to foods and meals (with EAT) and beverages (with DRINK). For this, we relied upon wordlists, by descending order of frequency, generated by Wordsmith™. We then extracted from those wordlists the top 20 such nouns occurring in these wordlists. Tables 8 and 9 summarize these results.

Table 8. Top 20 food-type nouns from object phrases of EAT in the sBNC and wBNC (most generic items in bold)

	EAT (sBNC)		EAT (wBNC sample)	
rank	top food types	N	top food types	N
1	***food***	**57**	***food***	**65**
2	***dinner***	**40**	***foods***	**29**
3	*meat*	28	*fish*	27
4	*cake*	22	*bread*	24
5	*bread*	21	*meat*	23
6	*chocolate*	21	***meals***	**17**
7	*fish*	21	***breakfast***	**16**
8	*cheese*	20	*cheese*	16
9	*chicken*	18	***lunch***	**14**
10	*chips*	18	***meal***	**14**
11	*fruit*	18	*cake*	13
12	*flies*	16	*chocolate*	12
13	***breakfast***	**15**	*cream*	12
14	***tea*** (meal sense)	**14**	*leaves*	10
15	*biscuits*	11	*fibre*	9
16	***meal***	**11**	*cakes*	8
17	*toast*	10	*rice*	8
18	*vegetables*	10	*sandwiches*	8
19	*cream*	9	***supper***	**8**
20	***lunch***	**9**	*vegetables*	8

In the case of the lemma EAT, one can observe something of the varied (and not entirely unhealthy) eating habits of the British, bearing in mind that the occurrence of so many *flies* in Table 8 is due to one particular repetitious classroom lesson. An interesting difference between spoken and written modalities is the occurrence of both *fish* and *chips* in the top 10 of the spoken corpus, whereas only *fish* occurs in the top 20 of the written. This could be a reflection of rather different preferences associated with informal and formal social settings. It can be seen that the most frequent object word in both the spoken and written corpora is *food*. In addition to the generic *food*, the top 20 lists for both corpora include names for meals, e.g., *breakfast*, *lunch*, *tea*, *supper*, *dinner*, as well as the words *meal* and *meals* themselves. It is commonplace in dictionaries to recognize a 'food' and 'meal' kind of understood object of intransitive EAT, corresponding to Huddleston and Pullum's (2002) categories of 'normal category indefinites' and 'specific category indefinites'. Our results show that these two categories are a feature of the *transitive* use of EAT as well. Intuition

tells us that EAT can occur with such object nouns, but only a corpus linguistic study as we have done tells us something about the robustness of this pattern. Note also the differing relative frequencies with which the main 'meal' words are mentioned in spoken and written corpora. In the sBNC, the relative frequency is *dinner* (40) > *breakfast* (15) > *tea* (14) > *lunch* (9), whereas in the sampled wBNC it is *breakfast* (16) > *lunch* (14) > *supper* (8). While we find *breakfast* > *lunch* in both corpora, the high frequency of *dinner* in the sBNC is noteworthy (*dinner* does not even appear in the top 20 food-type nouns of the sampled wBNC). Well over half of the *dinner* object phrases in the sBNC involve a possessive pronoun (*my dinner*, *your dinner*, etc.), typically used reflexively to refer back to the subject as in *I didn't eat my dinner*. We see here, perhaps, a subtle difference between spoken and written usage.

Table 9. Top 20 beverage-type nouns from object phrases of DRINK in the sBNC and wBNC (alcoholic items in bold)

	DRINK (sBNC)		DRINK (wBNC sample)	
rank	top beverage types	N	top beverage types	N
1	*tea*	46	*coffee*	43
2	*coffee*	41	*tea*	39
3	*milk*	23	***wine***	**33**
4	*water*	22	*water*	32
5	***wine***	**15**	***beer***	**26**
6	*drink*	10	***alcohol***	**18**
7	*juice*	9	*milk*	14
8	*orange*	9	*juice*	9
9	***beer***	**8**	***champagne***	**8**
10	*coke*	8	***brandy***	**6**
11	***alcohol***	**7**	*fruit*	6
12	***spirits***	**7**	***sherry***	**6**
13	***gin***	**6**	***whisky***	**6**
14	***sherry***	**6**	*blood*	5
15	***whisky***	**6**	***lager***	**4**
16	*drinks*	4	*mineral*	4
17	***ale***	**3**	***whiskey***	**4**
18	***brandy***	**3**	***ale***	**3**
19	***methylated***	**3**	***booze***	**3**
20	*pop*	3	*fluids*	3

With DRINK, we do not find an object noun *drink* or *beverage* occurring with the same kind of frequency as *food* does in the case of EAT. Neither

drink/*drinks* nor *beverage* occurs in the top 20 object nouns in the wBNC, for example. Instead, we find words with more specific kinds of meanings. Clearly, alcohol is a common type of object of transitive DRINK and not just a feature of the interpretation of intransitive DRINK. The occurrence of names for alcoholic beverages is striking, accounting for a clear majority of the top 20 beverage-type nouns as objects in both spoken and written corpora. Nevertheless, it is worth noting that *tea* and *coffee* are the most frequent in both corpora. *Tea* is the most common beverage-type noun in the sBNC, while *coffee* is the most common such noun in the wBNC, reflecting (as with *fish* and *chips* above) possible differences in preferences in informal versus formal settings. Ethnographically speaking, we also notice that this famously ale-drinking culture has discovered the grape: there are more instances of *wine* in the object phrase than *beer* in both corpora.

We find these results concerning the top 20 food and drink objects of some interest. Lexicographic practice typically identifies specialized intransitive uses of EAT and DRINK involving the specific interpretations of 'eat a meal' and 'drink alcohol', but omits any mention of these meanings with the transitive usage. This is understandable when a dictionary is intended to be used primarily to help users decode a particular usage of a verb. One might, for example, rely on the dictionary to 'fill in' an understood, but unexpressed object. As defensible as it may be for lexicographers to make inferences explicit in one case (the intransitive), but not the other (the transitive), this practice has the drawback of suggesting a difference between transitive and intransitive use when, as in this case, none exists. We stress, again, the virtue of corpus linguistic techniques for the descriptive linguist and lexicographer alike. By sampling thousands of instances of actual uses of an item, the full extent of inferences and collocational properties associated with a verb becomes apparent and the ensuing description becomes more observationally adequate.

Further differentiation of object nouns according to the inflected form of the verb yields additional information. Tables 10 and 11 provide a breakdown of the top 20 food-type nouns according by inflected form of EAT. The lists in these tables provide tantalizing glimpses into interactions between TAM marking and lexical properties of the objects. For example, while the superordinate term *food* is the most frequent word in almost all these lists, it is conspicuously absent with the simple past tense *ate* in the sBNC. Note also that *meat* is absent as an object of *ate* and *eaten* in both tables, both telic and highly episodic inflections of the verb EAT. These two object nouns, *food* and *meat*, occur most typically in 'habitual' contexts

(e.g., *The rich eat too much meat and suffer from chronic constipation,* from the sBNC). The absence of *food* and *meat* in these cases may be motivated by a disharmony between these words (and their habitual associations) and the simple past tense (with more 'past' and 'completed' associations).

Table 10. Top 20 food-like nouns with inflected forms of EAT in the sBNC

lexeme	head nouns in object phrase (N)
eat	***dinner (33), food (30****), meat (18), cheese (14), chips (13), bread (12), cake (12), fruit (11),* ***breakfast (11),*** *fish (11), chicken (10),* ***tea (10),*** *flies (9), biscuits (8), chocolate (8),* ***meal (8****), vegetables (8), apples (6), sandwiches (6), toast (6)*
eats	***food (8),*** *cheese (4), meat (4), fish (3), cake (2), flies (2), fruit (2) salads (2), sweets (2)*
eating	***food (14),*** *chocolate (7), cake (6),* ***dinner (5),*** *fish (5), meat (5),* ***supper (5), breakfast (4),*** *crisps (4),* ***lunch (4),*** *bread (3), chicken (3), cream (3), fruit (3),* ***meal (3), tea (3),*** *butter (2), chips (2), chocolates (2)*
ate	*flies (4), bread (3), chocolate (3), biscuit (2), cake (2), chips (2), cream (2), eggs (2), margarine (2), potatoes (2), pudding (2), stuffing (2), vegetables (2)*
eaten	***food (5),*** *toast (3), birds (2), bread (2), cheese (2),* ***dinner (2),*** *fish (2), hat (2) sausage (2)*

One particular sequence of specific words in our database that deserves comment is *something to eat*. This sequence is, in fact, the most frequent trigram which includes a form of EAT or DRINK in our database. We returned to the BNC, making use of the additional options in the BNCWeb application, to check on the statisitical significance of *something to eat*. We chose *to eat* as the node phrase and sought statistics on the word occurring immediately to the left. In the sBNC we found that *something to eat* occurred 53 times (mutual information score 6.65, Z-score 72.30); in the whole wBNC, *something to eat* occurred 153 times (mutual information score 7.48, Z-score 164.28). These scores indicate significant collocations in both corpora. Again, a corpus-based approach to language analysis can draw our attention to common usage, as opposed to the constructed examples of grammar books. Though *something to eat* is the most common trigram in our database, it is a usage of EAT which is often marginalized in

discussions of transitive verbs. Indeed, the infinitival complement of a noun is a frequently omitted construction type or category entry in reference grammars of English.

Table 11. Top 20 food-like nouns with inflected forms of EAT in the sampled wBNC

lexeme	head nouns in object phrase (N)
eat	***food (31)**, meat (16), bread (14), fish (14), **foods (13)**, **meals (8)**, cake (7), fruit (7), **breakfast (6)**, **lunch (6)**, **meal (6)**, cheese (5), cream (5), fibre (5), grass (5), salad (5), vegetables (5), cakes (4), chocolate (4), ice (4),*
eats	***food (4)**, fish (2), meat (2)*
eating	***food (17)**, **foods (14)**, cheese (7), **breakfast (5)**, chocolate (5), fish (5), fruit (5), **meals (5)**, sandwiches (5), bread (4), cream (4), fibre (4), animal (3), berries (3), cake (3), cereals (3), fat (3), heart (3), leaves (3), **meal (3)***
ate	***food (9)**, **lunch (6)**, fish (5), eggs (4), **meals (4)**, apple (3), beans (3), bread (3), **breakfast (3)**, cheese (3), chocolate (3), cream (3), rice (3), biscuits (2), cakes (2), cereal (2), chips (2), crisps (2), **dinner (2)**, ice (2)*
eaten	***food (4)**, **meal (3)**, bread (2), **breakfast (2)**, cake (2), **foods (2)***

6. Subjects

It is natural that there should be more focus on the nature of the object than the subject in discussions of transitivity. However, we are interested in gaining a better understanding of the whole transitive construction in English which includes both a subject and an object. We therefore examined properties of the subject phrases as well as the object phrases. In our database, a non-animate subject was extremely rare, though they did occur. Examples of inanimate subjects with EAT are given in (8). Newman (1997) discusses the metaphorical mappings that underlie these extensions and other ones based on EAT in English.

(8) Inanimate subjects (underlined) occurring with EAT

a. *A tall order, when <u>tennis time</u> eats into valuable study time*. (sBNC)

b. *...because a hangover had already eaten into his small reserves of patience and equanimity. (*wBNC)
c. *If your water is soft and acid, it will eat into the shell and dissolve it.* (wBNC)
d. *If she will be earning, that will eat into her profit.* (wBNC)
e. *Every mile of dual carriageway eats up twenty-six acres of countryside*. (sBNC)

We coded subject phrases by number and person for EAT and DRINK in both corpora. Figures 5 and 6 summarize the results based on counts of lemmatized EAT and DRINK. These results show parallel patterns for EAT and DRINK in each corpus (the sBNC and the sampled wBNC), though the difference between the spoken and written modalities is quite striking. In the spoken modality, the contour is defined by peaks at 1st singular, 2nd singular/plural, and 3rd singular, with a certain number of uninterpretable subjects as part of the corpus. In the written modality, on the other hand, 3rd singular, 3rd plural, and unspecified subjects predominate, with no uninterpretable subjects. These different distributions conform to some expected patterns, e.g., the high incidence of reference to speech act participants in the spoken language, at least in conversation. Conversely, there is a predominance of 'others', i.e., 3rd person forms, and unspecified subjects in the written corpus. While the overall trends evident in these figures may be well motivated, one cannot predict the specific distributions of individual verbs such as EAT and DRINK without an examination of a corpus. By the same token, one would need to carry out a comparable analysis of other verbs to be confident about the extent to which the profiles in Figures 5 and 6 (or even Figures 3 and 4) are replicated for other verbs.

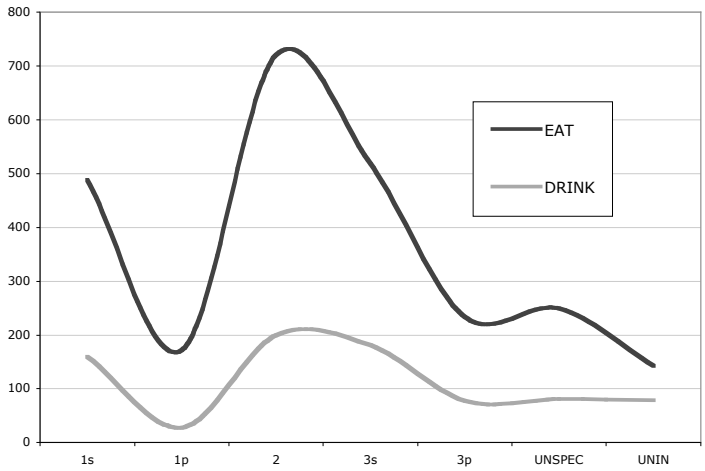


Figure 5. Raw frequencies of the lemmas EAT and DRINK by subject NP in the sBNC

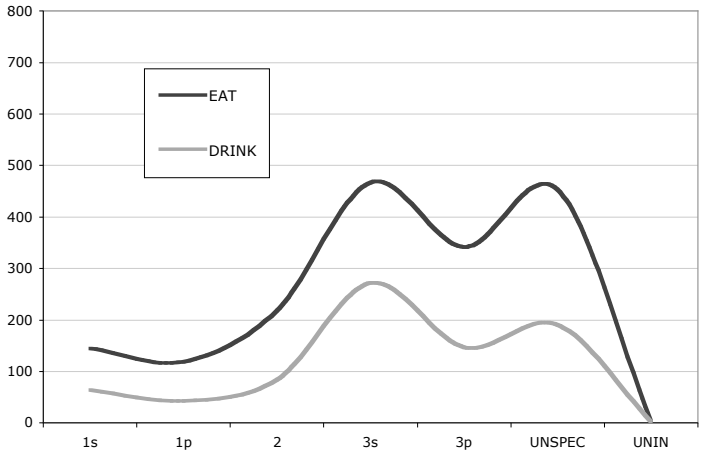


Figure 6. Raw frequencies of the lemmas EAT and DRINK by subject NP in the wBNC sample

7. Beyond (and below) the transitivity of EAT and DRINK

A couple of broad conclusions can be drawn from our corpus study into the transitivity alternations evinced by English EAT and DRINK. First of all, we completely concur with Hopper and Thompson's (1980) view that transitivity is scalar and that a verb's transitivity profile can vary (a variant of which was expressed in Rice 1987). However, is the phenomenon of transitivity the rightful place to start? We have found ample evidence that the presence or absence of an object phrase can vary by inflection and modality. Furthermore, the semantic properties of a verb and its overt or supressed arguments are construction-specific. Conversely, a verb's argument structure(s) should not be construed as hard-wired in the verbal lexicon, but as emergent from patterns of usage (which, needless to say, are genre- and modality-specific). This conclusion, self-evident to any linguist who works with a corpus, is only recently finding its way into theories of syntax, which have long been dominated by claims that context-free, language-wide, and universally inspired phrase structure patterns are the relevant unit of analysis. Where once truth was sought in the most generalized, category-based phenomena, we prefer to seek truth as it presents itself to us in the specifics of usage.

There is a growing realization amongst cognitively and functionally minded linguists that individual words, together with their co-occurring collocates, are not just a proper 'unit' of analysis, but represent a desirable descriptive and analytical starting point. That is to say, it is not just at the categorial level (N, NP, etc.) or the level of the lemma (EAT, DRINK) where we find patterns worthy of study. Words (*eat*, *drink*, etc.), together with their collocational forms, have become a focus of interest in a number of current approaches, notably Langacker's Cognitive Grammar (1987, 1991). It is an idea which also finds expression in Croft's Radical Construction Grammar (Croft 2001). Croft allows for specific constructions such as [*roasted* MEATNOUN] and [*toasted* BREADNOUN], alongside the superordinate construction [TRVERB-PASSPART NOUN]. Here, the specific words *roasted* and *toasted* help define two separate constructions, at one level of analysis. In essence, a verb's selectional restrictions, once exclusively relegated to the lexicon, are allowed to direct the syntax, at least in certain cognitively inspired theories of grammar. Moreover, Renouf and Sinclair (1991) have tracked the incidence of 'frames' such as *a(n) X of* (e.g., *a lot of, a kind of, an example of*); *too X to* (e.g., *too late to, too much to*); *many X of (many years of, many thousands of*) and show how the frame provides a significant context for the keyword X and vice versa (i.e.,

specific keywords dominate the frame). Similarly, Stefanowitsch and Gries' (2003) idea of a *collostruction*, understood as constructions in which particular lexical items play a key role, e.g., the [INTO-causative] construction or the [NP WAITING-TO-HAPPEN] construction, takes the study of grammar in similar directions.

In the same vein, Thompson and Hopper (2001: 44) recognize a crucial lexico-syntactic level of analysis which is built around a specific verb and its collocational forms. They write:

> ...among the things speakers know about verbs is the range of forms they collocate with according to the different senses they have...[T]he more different types of language speakers are exposed to and participate in, the wider the range of options for a given verb sense they are likely to have entered and stored... *[S]ome collocations involving specific verb senses develop lives of their own.* [Italics ours]

We would add that some collocations involving specific verbs *in specific inflections* develop lives of their own. It is not just certain verb lemmas which show an affinity for particular subject and TAM inflection or for realized or implied objects, but also the inflected forms themselves which do.

Some recent case studies of individual verbs in English have detailed such inflectional idiosyncrasy. Tao (2001, 2003) shows that the transitive lemma REMEMBER is overwhelmingly used, in the three spoken corpora he investigated, without an object complement, in the simple present tense, with first person singular (*I remember*) or null subjects (*remember?*), and at utterance boundaries. He concludes that the verb is well on its way to grammaticalizing into a discourse particle which regulates participant interaction in conversation and considers that a preoccupation with REMEMBER's argument structure and lexical meaning is misplaced. Scheibman (2001), in a study of informal conversation, found that 1st singular and 2nd singular subjects occur with particular verbs of cognition with a relative high frequency (*I guess*, *I don't know*, *you know*, *I mean*) reflecting the pragmatic value of such combinations in conversation. Scheibman (2001: 84) emphasizes the need to examine 'local' patterns in grammatical research and cautions against relying just on the superordinate grammatical categories (person, verb type, tense etc.). More recently, Newman and Rice (2004) describe the emergence of a *SIT around and...* construction where the meaning relates to futile, lazy, or otherwise unproductive activity rather than the posture of sitting. The presence of the

lexical item *around* is a crucial part of this construction. They also report on how the inflectional differences between the *–ing* and past tense forms in the pairs *sitting and.../sat and...*, *standing and.../stood and*, *lying and.../lay and...* profoundly influence the range of following verbal collocates. Finally, Rice and Newman (2004), in a study of aspectual uses of English prepositions, note that with the 'resumptive' construction *V on with*, just three collocate verbs (*get*, *carry*, and *go*) account for 90% of the 506 examples in the sBNC. They show, too, that the inflectional categories are distributed in construction-specific ways. Thus, the V in the 'continuous activity' *V on* construction occurs preferentially as a bare stem, whereas the V of the 'semeliterative' or 'corrective' *V over* construction occurs preferentially in the simple past.

The main message we want to impart is that inflected verb forms have their own semantic and constructional properties (hence, the reference to 'below transitivity' in the title of this section) and these merit serious descriptive and theoretical consideration. To that end, we propose the notion of an *inflectional island*, taking Tomasello's (1992) notion of a *verb island* a step further. He coined this term to describe the fact that morphosyntactic inflection tends to affect individual verbs in early child language and that syntactic development emerges from one verb to another and not across a lexical class as a whole. We use *inflectional island* in a similar fashion: syntactic/semantic properties tend to inhere in individual inflections of a verb in a register-specific manner. Furthermore, these properties may not extend across all the inflections to characterize the lemma as a whole. For us, the notion of a dictionary entry based on a lemma is still inadequate. Langacker's (1987: 63-76) dictum that grammar is a structured inventory of conventionalized units continues to provide a reliable and insightful way of conceptualizing language. The onus falls on us to identify and describe the level and nature of this conventionalization.

Appendix 1

Raw frequencies and percentages of person/number occurrences with the word form *eat* in our database.

S	*eat* O (s)		*eat* (s)		*eat* O (w)		*eat* (w)	
1s	291	20%	48	18%	49	8%	28	10%
1p	80	6%	38	14%	41	7%	33	12%
2	486	34%	67	25%	129	21%	55	21%
3s	214	15%	34	13%	101	17%	65	24%
3p	121	8%	44	16%	152	25%	40	15%
UNSPEC	176	12%	18	7%	133	22%	47	18%
UNIN	81	6%	20	7%	1	0%	0	0%
TOTALS	1449	100%	269	100%	606	100%	268	100%

Appendix 2

Raw frequencies and percentages of person/number occurrences with the word form *drink* in our database.

S	*drink* O (s)		*drink* (s)		*drink* O (w)		*drink* (w)	
1s	86	24%	25	20%	21	11%	13	11%
1p	14	4%	1	1%	12	6%	6	5%
2	123	35%	39	31%	42	22%	24	20%
3s	49	14%	25	20%	39	20%	35	29%
3p	27	8%	18	14%	29	15%	31	26%
UNSPEC	21	6%	5	4%	48	25%	12	10%
UNIN	35	10%	14	11%	2	1%	0	0%
TOTALS	355	100%	127	100%	193	100%	121	100%

Notes

1 An earlier version of this paper was presented at the theme session on *Language Between Text and Mind: The Use of Corpora in Cognitive Linguistics* at the 8th International Cognitive Linguistics Conference in July, 2003. We would like to thank our research assistants, Hui Yin and Hideyuki Sugiura, who carried out preliminary coding of our corpus search results. Thanks also to the editors and reviewers of this volume who provided helpful feedback.

2 We use small capital letters to denote a lemma which subsumes all the inflected forms, e.g., EAT, and italics to denote a particular word form or lexical item. Thus, EAT subsumes *eat*, *eats*, *eating*, *ate*, *eaten*.

3 For Levin (1993: 42, 213-214), EAT and DRINK exhaust their particular sub-class of what she calls *Verbs of Ingesting*, an exclusivity which makes them all the more intriguing.

4 Note that Van Valin and LaPolla's (1997) distinction between non-referential objects (*He ate spaghetti in ten minutes*) and referential objects (*He ate the plate of spaghetti in ten minutes)* is construable as a special case of parameter J, individuation of the object, in Hopper and Thompson's (1980) model of Transitivity.

5 We reserve the term 'modality' for spoken versus written modes of communication, whereas we take 'genre' to mean a register difference within a modality, e.g., spontaneous conversation, story-telling, or ceremonial language.

6 Despite our qualms, this grammar is a prodigious achievement nevertheless.

7 We have opted to use lines instead of bars to express quantities in many of the following figures. We felt a need to collapse information between corpus, verb, and transitivity class, as well as across inflectional category. We do not intend for these line-based figures to give the impression of continuous functions across what are obviously discrete categories. However, the lines constitute a distributional 'profile' which is easier to assess and compare than what would otherwise be a proliferation of individual bars. We use the following abbreviations in these figures: 1 = 1st person, 2 = 2nd person, 3 = 3rd person, s = singular, p = plural, UNSPEC = unspecified, UNIN = uninterpretable.

References

Alm-Arvius, Christina
1993 *The English Verb* SEE*: A Study in Multiple Meaning*. Göteburg: Acta Universitatis Gothoburgensis.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan
2000 *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.

Cooper, William E., and John R. Ross
1975 Word order. In *Papers from the Parasession on Functionalism,* R. E. Grossman, L. J. San, and T. J. Vance (eds.), 63-111. Chicago: Chicago Linguistic Society.

Croft, William
2001 *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Hopper, Paul J., and Sandra A. Thompson
1980 Transitivity in grammar and discourse. *Language* 56: 251-299.

Huddleston, Rodney
1988 *English Grammar: An Outline*. Cambridge: Cambridge University Press.

Huddleston, Rodney, and Geoffrey K. Pullum (eds.)
2002 *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

Kilgarriff, Adam
2001 Comparing corpora. *International Journal of Corpus Linguistics* 6: 97-133.

Langacker, Ronald W.
1987 *Foundations of Cognitive Grammar*, Vol. I, *Theoretical Prerequisites*. Stanford: Stanford University Press.
1991 *Foundations of Cognitive Grammar*, Vol. II, *Descriptive Application*. Stanford: Stanford University Press.
2001 A dynamic usage-based model. In *Usage-based Models of Language,* Michael Barlow and Suzanne Kemmer (eds.), 1-63. Stanford: CSLI Publications.

Levin, Beth
1993 *English Verb Classes and Alternations*. Chicago: University of Chicago Press.

Lichtenberk, Frantisek
1991 Semantic changes and heterosemy in grammaticalization. *Language* 67: 475-509.

Newman, John
1996 *Give: A Cognitive Linguistic Study*. Berlin/New York: Mouton de Gruyter.

1997 Eating and drinking as sources of metaphor in English. *Cuadernos de Filología Inglesa* (Special volume on Cognitive Linguistics) 6: 213-231.

2004 Motivating the uses of basic verbs: Linguistic and extra-linguistic considerations. In *Motivation in Grammar*, Günther Radden and Klaus-Uwe Panther (eds.), 193-218. Berlin/New York: Mouton de Gruyter.

Newman, John (ed.)

1998 *The Linguistics of Giving*. Amsterdam/Philadelphia: John Benjamins.

2002 *The Linguistics of Sitting, Standing, and Lying*. Amsterdam/Philadelphia: John Benjamins.

Newman, John, and Sally Rice

2001 English SIT, STAND, and LIE in small and large corpora. *ICAME Journal* 25: 109-133.

2004 Patterns of usage for English SIT, STAND, and LIE: A cognitively-inspired exploration in corpus linguistics. *Cognitive Linguistics* 15: 351-396.

Norvig, Peter, and George Lakoff

1987 Taking: A study in Lexical Network Theory. In *BLS* 13: 195–206.

Pauwels, Paul, and Anne-Marie Simon-Vandenbergen

1995 Body parts in linguistic action: Underlying schemata and value judgements. In *By Word of Mouth: Metaphor, Metonymy and Linguistic Action in a Cognitive Perspective*, Louis Goossens, Paul Pauwels, Brygda Rudzka-Ostyn, Anne-Marie Simon-Vandenbergen, and Johan Vanparys (eds.), 35-69. Amsterdam/Philadelphia: Benjamins.

Payne, Doris L.

2001 The explanation of argument structure: Lexicalization or discourse use? *Proceedings of the Seoul International Conference on Discourse and Cognitive Linguistics*, 532-556. Seoul: Korean Society of Discourse and Cognition.

Poutsma, Hendrik

1926 *A Grammar of Late Modern English, Part II: The Parts of Speech, Section II: The Verb and the Particles*. Groningen: P. Noordhoff.

Radden, Günther

1996 Motion metaphorized: The case of 'coming' and 'going'. In *Cognitive Linguistics in the Redwoods: The Expansion of a New Paradigm in Linguistics*, Eugene H. Casad (ed.), 423–458, Berlin/New York: Mouton de Gruyter.

Renouf, Antoinette, and John M. Sinclair

1991 Collocational frameworks in English. In *English Corpus Linguistics: Studies in Honour of Jan Svartvik,* Karin Aijmer and Bengt Altenberg (eds.), 128-143. London/New York: Longman.

Rice, Sally

1987 Towards a cognitive model of transitivity. Ph.D. diss., Department of Linguistics, University of California, San Diego.

1988 Unlikely lexical entries. *BLS* 14: 202-212.

Rice, Sally, and John Newman

2004 Aspect in the making: A corpus analysis of English aspect-marking prepositions. In *Language, Culture, and Mind*, Michel Achard and Suzanne Kemmer (eds.), 313-326. Stanford: CSLI Publications.

Scheibman, Joanne

2001 Local patterns of subjectivity in person and verb type in American English conversation. In *Frequency and the Emergence of Linguistic Structure*, Joan Bybee and Paul Hopper (eds.), 61-89. Amsterdam/Philadelphia: John Benjamins

Serra Borneto, Carlos

1996 *Liegen* and *stehen* in German: A study in horizontality and verticality. In *Cognitive Linguistics in the Redwoods: The Expansion of a New Paradigm in Linguistics*, Eugene H. Casad (ed.), 459–505. Berlin/New York: Mouton de Gruyter.

Shen, Ya-Ming

1996 The semantics of the Chinese verb 'come'. In *Cognitive Linguistics in the Redwoods: The Expansion of a New Paradigm in Linguistics*, Eugene H. Casad (ed.), 507–540. Berlin/New York: Mouton de Gruyter.

Stefanowitsch, Anatol, and Stefan Th. Gries

2003 Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8: 209-243.

Sweetser, Eve

1990 *From Etymology to Pragmatics*. Cambridge: Cambridge University Press.

Tao, Hongyin

2001 Discovering the usual with corpora: The case of *remember*. In *Corpus Linguistics in North America: Selections from the 1999 Symposium*, Rita C. Simpson and John M. Swales (eds.), 116-144. Ann Arbor: University of Michigan Press.

2003 A usage-based approach to argument structure. *International Journal of Corpus Linguistics* 8: 75-95.

Thompson, Sandra A., and Paul J. Hopper

2001 Transitivity, clause structure, and argument structure: Evidence from conversation. In *Frequency and the Emergence of*

Linguistic Structure, Joan Bybee and Paul Hopper (eds.), 27-60. Amsterdam/Philadelphia: John Benjamins

Tomasello, Michael

1992 *First Verbs: A Case Study of Early Grammatical Development*. Cambridge: Cambridge University Press.

Van Valin, Jr., Robert D., and Randy J. LaPolla

1997 *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.

Wierzbicka, Anna

1988 Why you can have a Drink when you can't have an Eat. *The Semantics of Grammar*, Anna Wierzbicka, 293–357. Amsterdam/Philadelphia: Benjamins.