

Article drop in English headlines

Andrew Weir (832110)

August 2009

Submitted in partial fulfilment of the requirements for the degree of MA Syntax

University College London

Faculty of Life Sciences

Division of Psychology and Language Sciences

Research Department of Linguistics

Supervisor: Dr Klaus Abels

Contents

1	Introduction	4
2	Literature review	6
2.1	Straumann (1935)	6
2.2	Mårdh (1980)	6
2.3	Simon-Vandenberg (1981)	7
2.4	De Lange (2004)	7
3	Data	9
3.1	Types of headline	9
3.2	Distribution	9
3.3	Tasks	12
4	Pragmatics	14
5	Phonology	15
5.1	Generalised article drop	15
5.2	Phonology and string-initial ban on <i>a</i>	18
6	Semantic accounts	19
6.1	De Lange (2004)	19
6.2	Baauw et al. (2002)	20
7	Syntax	23
7.1	Truncation	23
7.2	High determiners	24
7.3	Details of a truncation analysis	24
7.4	Problems	26
7.5	Possible further work	28
8	Other forms of “block language”	29
8.1	Acquisition	29
8.2	Other characteristics of block language	30
9	Conclusion	31
	References	32

Acknowledgements

I would like to thank my supervisor, Klaus Abels, for his invaluable help, comments, suggestions, and criticisms in the course of writing this dissertation.

I would also like to thank (in alphabetical order) Amélie, Natalia, Oliwia, Sam, and Yiannis, for helping me keep body and soul together throughout the year.

Any omissions or errors of course remain my own.

Note

While many of the headlines presented as examples in this dissertation are genuine headlines, referenced to particular newspapers or other media, some of the examples have been invented in order to make particular linguistic points. Neither any created examples nor the choice of particular genuine examples are intended to cast aspersions on real persons living or dead, or to reflect any particular strand of political opinion.

1 Introduction

This dissertation will investigate the language used in headlines in English-language newspapers. In particular it will examine the phenomenon of article drop in headlines.

Article drop — and dropping of other elements, such as copulas — is a ubiquitous feature of headlines. Consider the below, headlines taken from the *Guardian* website <http://www.guardian.co.uk/> on 18 July 2009. Elements required in ‘standard’ written English, but which are in fact not present in the headlines, are shown in ⟨angle brackets⟩.

- (1) (a) ⟨A⟩ British first world war veteran dies at 113
- (b) China quarantines ⟨a⟩ UK school group
- (c) Purnell: I lost faith in ⟨the⟩ PM months ago
- (d) ⟨A⟩ Malaysian ⟨is⟩ suspected in ⟨the⟩ Jakarta blasts
- (e) ⟨The⟩ Hidden persecution of Falun Gong

Initially we may suspect that the register that includes headlines — or “headlinese”, as Mårdh (1980) terms it — just does not permit articles at all. But this is not the case, as the following headlines show (emphasis mine):

- (2) (a) National Express line is **an** ‘accident waiting to happen’ (*The Guardian*, 7 July 2009, p. 26)
- (b) Backpacker takes **a** well-trodden path to sell survival story (*The Times*, 17 July 2009)

Certain distributions of articles in headlines appear to be ill-formed, however. Intuitionistically, headlines of the form shown in (3) are ungrammatical:

- (3) *A man bites dog

In this dissertation I will argue that there are definite grammatical constraints on the distribution of articles in headlines. I will attempt to identify what precisely these constraints are, investigating the possibilities that they are phonological, semantic, syntactic, or pragmatic in nature. Drawing on the data provided by the corpus investigation of Mårdh (1980), as well as intuitionistic data, I will suggest that article drop is a generally free procedure, sanctioned by an ‘extra-syntactic’ introduction of the function of the article into the semantics, as suggested by Baauw et al. (2002). There is, however, a further restriction on article drop: headlines may not begin with the article *a*. I explore the possibility of explaining this using a syntactic truncation analysis, following Rizzi (1994) and Haegeman (1997). While I argue that such an explanation would be satisfying, and would provide a way of unifying the case of headlinese with other cases of dropped

elements in reduced written registers (such as subject drop in diaries), the concrete attempt at providing such an analysis has too many problems to be currently workable. However, I hope that future work will go some way towards resolving this problem and explaining the constraints on article drop in headlines.

2 Literature review

While the matter of article drop in headlines has only rarely been the subject of study in the framework of generativist linguistics, it has occasionally been studied in works of a more general linguistic character. Here I will review four texts that cover article drop: Straumann (1935), Mårdh (1980), Simon-Vandenberg (1981), and de Lange (2004).

2.1 Straumann (1935)

Straumann (1935) is the first work to attempt to systematically analyse the language of English headlines. Although the book obviously predates the advent of generativist linguistics, the study is in-depth and detailed. It is also one of the first works to point out that the language of headlines, in common with other 'reduced' registers of English — which Straumann names "block language" — has its own syntax which, although clearly related to the syntax of the core grammar of spoken English, differs from it substantially, and in a way which is difficult to understand simply in terms of reduction of the possibilities offered by the core language ("block-language cannot in any case be comprehended by means of the traditional categories of grammar", p. 21). Straumann's study can therefore be seen as an early precursor of the notion that the 'reduced' written register merits a particular linguistic study, and can shed light on other areas of general linguistics.

Unfortunately, in terms of the present study, while Straumann covers many aspects of headlines in great detail, the status of articles is not one of them. In fact, Straumann contends that articles are specifically not worth studying, as he sees no systematic pattern governing their presence or absence: "[i]n headlines the treatment of the article is so absolutely arbitrary, i.e a matter of mere typographical caprice, that [...] it must be left out of consideration" (p. 51).

As we will see later in the present work (and as stated in the work of other authors), this contention is wrong. There are definite grammatical, non-pragmatic conditions on the presence or absence of articles in headlines, which cannot be reduced to "typographical caprice" or such pragmatic restrictions. So, while Straumann's study provides an interesting basis for the study of reduced register/block-language in general, it is not directly relevant to a study of article drop in headlines and reduced register.

2.2 Mårdh (1980)

Mårdh (1980) is a monograph investigating the grammar of headlines. It does not adopt a particular theoretical viewpoint, and certainly not a generativist one. Rather, it is an empirical study and presentation of the data, using headlines from the (London) *Times* and *Daily Mirror* in the 1970s to note peculiarities of headlines.

It dedicates an entire section (pp. 113–157) to analysis of the article, and the fact that it is often missing from headlines. While the study does not attempt to provide a theoretical analysis of the phenomenon of article drop, it does make a number of interesting empirical observations, the most notable of which for the present work is that noun phrases with an article do not appear to precede noun phrases without an article. Further analysis of the data that Mårdh provides shall be given in the Data section, below.

2.3 Simon-Vandenberg (1981)

Simon-Vandenberg (1981) is a study of the grammar of headlines in the (London) *Times* from 1870 to 1970. The study is in the tradition of functional grammar. As such, its insights are not directly transferable to the present work, which places itself in the generative tradition. Simon-Vandenberg does devote a section (pp. 276–84) to the ellipsis of the article; however, as the work is functionalist, it does not attempt to provide an explanation in ‘pure’ syntactic terms for the distribution of articles in headlines. Rather, it tries to find what might be termed ‘semantic’ or ‘pragmatic’ rationales for article drop, based on the information communicated by the headline; whether information is ‘new’ or ‘old’. As I will argue later in this dissertation, I contend that while there is a large semantic component to the explanation of article ellipsis in headlines, there are also restrictions which are not semantic in nature. As such I do not support the thesis of Simon-Vandenberg that semantics, or a functional explanation, alone can explain article ellipsis in headlines.

2.4 De Lange (2004)

De Lange (2004) is primarily a study of article drop in child language, rather than in headlines. Furthermore, de Lange’s study analyses article drop in Dutch and Italian child language, rather than English. As such it may not seem immediately relevant. However, de Lange draws parallels between child language and headlines which he suggests can provide a universal explanation for article drop cross-linguistically, both in headlines and in child language.

De Lange’s argument is based on information theory (Shannon 1948). De Lange notes (pp. 115f.) that, while Dutch has only three articles (*de*, definite, masculine or feminine gender; *het*, definite, neuter gender; *een*, indefinite, any gender), Italian has several. I will not list all the Italian possibilities here; the important point is that Italian articles have a full paradigm across gender, definiteness (including partitive articles as well as definite and indefinite), and plurality, showing very little syncretism. Furthermore, there are allophonic variants of certain articles in Italian. De Lange argues that Italian articles, as they contain more distinct grammatical and phonological features than their

counterparts in Dutch do, embody more information (in the technical sense of Shannon 1948) than Dutch articles. For example, in Dutch, information about the gender of a noun is not transmitted, at least in the indefinite case (as gender is not morphologically encoded on either the noun or the article); in Italian, it is, through the (overt morphological) gender feature on the article.

The de Lange paper establishes that article drop is considerably less frequent in Italian than in Dutch, in both headlines and in child speech. De Lange argues that this is because of considerations of transmitting the maximum of information. In headlines, de Lange argues, editors wish to convey as much information as possible. Removing a low-information element increases the overall information density of a headline. In Dutch, articles are low-information elements, and so can felicitously be dropped from headlines. In Italian, articles are high-information elements, and so are resistant to deletion from headlines. In child speech, a similar phenomenon occurs. De Lange argues, based on work by Kostić (2004) that low-information elements do not 'activate' a child's processing ability as strongly as high-information elements do. As such, in generation of speech, low-information elements are particularly likely to be dropped. De Lange claims that the empirical evidence concerning article drop by Dutch- and Italian-speaking infants bears this out.¹

So far, so good. However de Lange also makes a claim which is directly relevant to a study of article drop in English headlines. De Lange refers to a generalisation first noted by Mårdh (1980) (as mentioned above) and developed by Stowell (1999): that, in English headlines, a DP with an article apparently cannot precede a DP that has been subject to article drop (the **A man bites dog* example). De Lange argues that his information-theoretical framework can explain this. De Lange argues (p. 117) that the sentence-initial position is the canonical position for 'old' information. As such, this discourse feature of a DP in sentence-initial position is encoded by virtue of that position, and does not need to be encoded again by an article. This results in articles on DPs in sentence-initial position bearing a lower information load (as they do not need to encode the discourse feature 'old information') than articles in a later position, and thus they are more susceptible to drop. Articles in a later position, however, are required to encode discourse features, and so bear a relatively high information load and are less susceptible to drop.

I believe that this argument cannot satisfactorily explain the data. One problem that can be highlighted immediately is that the sentence-initial position is clearly not a position for 'old' information; in the sentence *A man bites a dog*, for example, *a man* is not old information. Further problems with this information-theoretic approach will be explored in more detail later in this dissertation.

¹I am here slightly glossing over the technical details of this proposal, both for ease of exposition, and because I do not have access to Kostić (2004).

3 Data

In this section, I will set out data that is relevant to the question of article drop in English headlines. The data is intuitionistic, but an effort has been made to search through the online Nexis database of headlines in order to confirm that the intuitions correspond to empirical fact. I have not, however, performed any in-depth investigation of the Nexis database as a corpus.

3.1 Types of headline

Several types of headline do not show the phenomenon of article drop, and as such have been excluded from investigation. In particular, article drop appears to only occur in headlines for news articles. Headlines for commentary or editorial pieces do not appear to show article drop.² Subheads — smaller headlines accompanying the main headline — also appear not to show article drop.

Furthermore, quotes — i.e. language attributed to a speaker other than the newspaper — in headlines have a lesser tendency to drop articles (see Mårdh 1980:131f.). The tendency decreases further if the quote is direct, and appears to be completely absent (i.e. no articles are dropped at all) if the quote is enclosed in quote marks. Presumably this stems from a desire to keep a direct quote unchanged from the original.

All of these types of headline — commentary, subhead, quoted language — will therefore be disregarded from consideration.

3.2 Distribution

Articles are very rarely obligatory in headlines. In general, dropping all articles in a headline delivers a grammatical and ‘idiomatic’ result (again, ⟨angle brackets⟩ indicate deletion):

- (4) ⟨A⟩ Man bites ⟨a⟩ dog
- (5) ⟨A⟩ Disaster as ⟨a⟩ hurricane strikes ⟨the⟩ south coast of ⟨the⟩ United States
- (6) ⟨The⟩ TUC pledges moderation but ⟨the⟩ threat of ⟨a⟩ clash in engineering grows (Mårdh 1980:121)³

There are certain exceptions to this in specific circumstances. Mårdh (1980:153) notes, for example, that the article *a* used distributively, synonymously with *per* (e.g. *fifty miles an hour*) is never omitted. Articles in apposition structures are also rarely dropped, unless a comma intervenes between the two nouns in apposition

²In fact, they do not seem to show any differences with “normal” written English at all; none of the dropping phenomena appear to be present.

³Originally from the *Times*, 28 March 1974.

(1980:150–1); the same is true for structures where the copula is dropped. Mårdh gives as examples “NEVER AGAIN! SAYS SALLY **THE** STREAKER”⁴, for the apposition case, and “Purged transcripts a bestseller”⁵ for the copula drop case (emphasis mine in both cases).

Articles may in theory not be dropped at all. Just as articles are not obligatory, article drop is not obligatory either; in principle all the article-ful ‘translations’ of headlines given above could be used as headlines (although they rarely are). The interesting data is the possibility of some, but not all, articles being dropped in a given headline. Here we find the key contrast that indicates that article drop is not simply a free process, ‘optionally drop articles’. There is some grammatical conditioning at work here. Compare the two headlines below.

(7) Man bites a dog

(8) *A man bites dog

One of the tasks of the present work will be to try to offer an account for why the first of these is a grammatical headline, but the second is not permitted. This intuition is backed up by the empirical study of Mårdh (1980), who notes (pp. 141–3) that noun phrases with articles do not, in the corpus of headlines that she studies, precede noun phrases without articles. Examples like (9 a, b) are attested in her study, but examples like (9 c) are not.

- (9) (a) Irish Cabinet meets as city is declared a disaster area (Mårdh 1980:145)⁶
 (b) National Express line is an ‘accident waiting to happen’ (*The Guardian*, 7th July 2009, p. 26)
 (c) Irish Cabinet meets as the city is declared disaster area

Mårdh argues from the basis of this data that the relevant constraint is that noun phrases with an article present cannot linearly precede noun phrases with an elided⁷ article. I argue, however, that this is not the case. Intuitionistically, I claim that 9 c is perfectly grammatical, despite the fact that it is not attested in Mårdh’s study. Consider such examples as the below:

(10) Prime Minister sacks the Cabinet in new twist to expenses scandal

(11) Beijing hands over the flag to new organising committee

Although these examples are invented, not attested, I argue that they are unproblematic in headlines.

⁴Originally from the *Daily Mirror*, 19 March 1974.

⁵Originally from the *Times*, 2 May 1974.

⁶Originally from the *Times*, 18 May 1974.

⁷I use this term not to suggest that the process is necessarily one of ‘elision’, but simply to distinguish the case of article drop from that of noun phrases that would normally not take an article anyway, such as proper nouns, mass nouns, or plural count nouns.

I also reject another generalisation suggested in the literature: that is, the suggestion made by Stowell (1999), that the restriction on article drop is actually one of c-command; no article-less DP may be c-commanded by an article-ful DP. This generalisation is certainly tempting: it covers the alternation *Man bites a dog*/**A man bites dog*, mentioned above. It also explains examples (10, 11). If we assume that the constituents *in new twist to expenses scandal* and *to new organising committee* are adjuncts, positioned higher in the syntactic tree than the sentential objects *the Cabinet* and *the flag*,⁸ then those objects do not c-command the adjuncts, and so article drop in the adjuncts is licit.

A formulation in terms of c-command also explains certain “exceptions” Mårdh (1980) cites to the rule that she formulates. Consider, for example, one of Mårdh’s “exceptions” (p. 146):

(12) Mr Jo Grimond breaks an arm in Edinburgh fall

In (12), the adjunct *in Edinburgh fall* is (by assumption) positioned as an adjunct to the verb phrase, and is not c-commanded by the object *an arm*. There is therefore no violation of the c-command constraint.

However, it is not altogether obvious that a c-command generalisation is correct. Compare the pair below:

- (13) (a) Senior civil servant gave a secretary bribe
 (b) Senior civil servant gave secretary a bribe

Both of these sentences appear to my intuitions to be grammatical in headlines (or, at least, they are much less strikingly ungrammatical than the case of e.g. **A man bites dog*). Now (depending on one’s preferred theory of double-object verbs like *give*), in (13 a, b) either the *bribe*-DP c-commands the *secretary*-DP or the *secretary*-DP c-commands the *bribe*-DP. On some theories, both DPs c-command each other. But on no-one’s theory of double-object verbs does *neither* DP c-command the other. However, in order for both of these sentences to be grammatical, this is exactly what would have to be true, if the principle that an article-ful DP may not c-command an article-less DP is to be maintained.

If these generalisations are inaccurate, what is the correct generalisation? One generalisation that could be considered is that there is a restriction on matrix subjects: it is matrix subjects that must be without articles in headlines. In an earlier draft of this dissertation, this was my proposal; I considered that this would explain both the *Man bites a dog*/**A man bites dog* generalisation, as well as the cases in (13 a, b). What is *not* grammatical is the sentence in (14):

⁸This is, however, not an uncontroversial assumption; proponents of ‘rightward is downward’ theories of syntactic structure would presumably not want to suppose that adjuncts c-command objects. However, I will shortly argue that this is a moot point, as the relevant generalisation is in fact not one of c-command.

(14) *A senior civil servant gave secretary bribe

If the restriction on article drop is that matrix subjects must be article-less, then this would predict the ungrammaticality of (14). However, on reflection, I argue that this cannot be the correct generalisation. Consider the below sentences:

- (15) (a) Man bit dog, source claims
 (b) Man bit dog, a source claims

Both of these headlines seem to be grammatical. Note, however, that the matrix subject in both (15 a) and (15 b) is *source*. It does not appear to make a difference whether the matrix subject has an article or not. Now consider the below:

- (16) (a) Source claims that a man bit dog
 (b) *A man bit dog, source claims

All that is different between (16 a) and (16 b) is the placement of the quoted material *a man bit dog*. However, this difference appears to make a difference to the grammaticality of the headlines. The headline whose *string* begins with *a* appears to be ungrammatical. This constraint does not, however, seem to extend to the article *the*; consider the examples below:

- (17) (a) The Minister of Defence sacks senior civil servant
 (b) The Minister of Defence gave secretary bribe
 (c) The Minister of Defence gave secretary bribe, source claims

I suggest that all of the above sentences are grammatical in headlines. I propose, therefore, that the correct statement of the generalisation governing articles in headlines is as follows:

- (18) Articles can in general be optionally dropped in headlines; and the article *a* must not appear string-initially in headlines.

3.3 Tasks

Having established the generalisation in (18), our task now is to try to establish an analysis explaining what linguistic principles cause the generalisation to be true. Failing that, we can at least try to establish in what domain a possible analysis could lie (that is, whether the phenomenon is phonological, semantic, syntactic...).

There appear to be two problems to be tackled: firstly, why is article drop permitted at all in headlines (and apparently completely freely); and secondly, why must strings in headlines not start with the article *a*. Ideally an analysis would be found that could explain both of these, with an explanation for one

falling out from the principles that govern the other. However, it may also be possible that the two phenomena are entirely separate and require separate explanations. In the following sections of this dissertation, I will consider what domain the phenomenon of article drop in headlines belongs to, and propose a number of analyses which may be considered, but which in fact do not explain all the data. Unfortunately, due to constraints of time, that is as far as this dissertation can go in considering the matter; a definitive analysis of article drop in headlines will have to await further work.

4 Pragmatics

A pragmatic account of article drop would suggest that articles are dropped purely due to external, non-linguistic factors. For example, headlines often have to fit into a very restricted space, only a few characters long. If articles — as well as other elements, such as copulas or conjunctions⁹ — are optional, this allows the sub-editor compiling the headlines a certain flexibility to reduce the length of a headline if he or she needs to in order to fit it into the space available. Mårdh (1980:126ff.) discusses this possibility, among various other typographical motivations (such as maintaining an even line length across different lines) for dropping elements. A pragmatic account of article drop (and the dropping of other elements such as copulas) would argue that elements such as articles, whose semantic content is relatively low and which can generally be gleaned from context, are dropped solely in order to help the sub-editor.

This argument is similar to one which could be used to describe, for example, the grammar of note-taking, where the requirement for speed leads to the dropping of several elements that are perceived to be unnecessary. (See, for example, Janda 1985 for discussion of note-taking.)

While it is certainly true that the ‘flexibility’ of headlines does help the sub-editor, I maintain that this view of article drop ‘puts the cart before the horse’. The linguistic possibility of article drop comes first, which aids the sub-editor; but the desire for flexibility was not the *motivation* for the phenomenon of subject drop.

I argue this chiefly because of the below contrast:

- (19) (a) MAN BITES A DOG
 (b) A MAN BITES DOG

Both of the above are exactly the same length. And yet, as discussed in the ‘data’ section above, the most interesting fact about article drop in headlines is precisely that (19 a) is grammatical, while (19 b) is not. If the explanation for article drop were solely pragmatic and typographical, then we should not see a contrast between two headlines which are exactly the same length; and yet we do. It may be that the phenomenon of article drop somehow *originated* out of pragmatic motives, and that may be a reason for its continued existence in headlines; but a pragmatic account cannot provide an explanation for the distribution of the phenomenon.

⁹Conjunctions are not usually dropped in British headlines, but are commonly dropped in US headlines and replaced with a comma; e.g. *Barak, U.S. Envoy Discuss Settlements* (Washington Post, 3 July 2009)

5 Phonology

As headlines is a written phenomenon, it is not immediately obvious to what extent ‘phonology’ can be used as an explanation when one is examining written language. However, at least arguably, if evidence can be found that a plausible grammar of (spoken) language can be constructed which generates all and only the acceptable sentences of headlines, and that that grammar relies on phonological processes to generate the result (to drop the articles), then we can suggest that this is evidence for some form of phonological solution to the question of article drop in headlines. We do not necessarily need to consider the question of how or why a phonological component is present in written language; this is an interesting question, but an in-depth treatment of it is outside the scope of this dissertation. To that end, when considering phonological explanations of article drop, I will review arguments that suggest that infant speech drops articles because of phonological or metrical factors, and will consider whether such arguments can be generalised to the case of written English in headlines. Firstly, I will consider whether phonological arguments can account for article drop in the general case, and then I shall consider whether phonology can explain the apparent ban on string-initial *a*.

5.1 Generalised article drop

In an article by Gerken (1991), the argument is advanced that the phenomenon of dropped elements in infant speech in English is conditioned metrically. Whether an element (or, rather, syllable) can be dropped is dependent on its metrical position and metrical stress. Specifically, Gerken argues that the first (unstressed) position in an iambic foot is a position that allows drop of the syllable hosted there. The argument is reinforced by the work of McGregor and Leonard (1994) (who specifically discuss article drop in this context) in the case of infants with Specific Language Impairment. (See, however, Waller (1997) for some strong criticism of Gerken’s account.) I have previously argued (Weir 2009) that this argument can be recast for the adult spoken language in terms of intonational phrases; the first syllable in an intonational phrase can optionally be dropped. This is the explanation for ‘subject pronoun drop’, as well as many other types of drop, in the adult language, as exemplified below (Weir 2009’s (31), slightly adapted: ⟨angle brackets⟩ demarcate dropped syllables):

- (20) (a) ⟨I⟩ don’t think so.
(b) ⟨I’m⟩ going to the cinema.
(c) ⟨Are⟩ you going to the cinema?
(d) ⟨The⟩ man over there seems to think so.

Obviously it is this last example, the dropping of an article, which most concerns us here. Can we generalise the phonological accounts of Gerken and Weir in explaining dropped elements in infant and adult colloquial English, respectively, to the general case of article drop in headlines?

Initially, Weir (2009)'s argument concerning intonational phrase boundaries may seem attractive, particularly insofar as it would correctly rule out **A man bites a dog*; the dropped article is not initial in an intonational phrase. However, we can quite quickly determine that this argument does not generalise to all headlines, just by looking at the 'canonical' example set out in (21):

(21) ||⟨A⟩ Man bites ⟨a⟩ dog||

As indicated by the || delimiters, this headline (if one was to imagine reading it out) resides entirely within one intonational phrase (with phrasal stress falling on *dog*). While Weir's explanation allows for the deletion of the first article, it does not allow deletion of the second — which is nevertheless grammatical.

Gerken's proposal falls foul of a similar problem. Consider the foot structure in (22) below (constructed following Gerken's rules for foot structure (1991:437), and in particular, following the rule that feet are created in a maximally binary fashion from left to right:)

(22) (a MAN) (BITES a) (DOG)

Following Gerken, let us assume that weak initial syllables in iambs are candidates for drop. Then, from (22), we can certainly derive the grammaticality of *Man bites a dog*, and the ungrammaticality of **A man bites dog*. This follows from the fact that the second article in (22) is not in an iamb, but rather a trochee, which is argued by Gerken not to be a dropping environment. However, this very fact means that Gerken's analysis cannot deliver the 'canonical' *Man bites dog*; if the second article in (22) cannot be dropped, then there is no way of getting there.

We might try a variation on Gerken's analysis, adapted for headlines: to say that articles in 'weak' position, unstressed in an iamb, *must* drop; articles in other positions may or may not drop, optionally. Headlines with all articles present, such as *A man bites a dog*, could be argued to be a result of code-switching out of headlines 'proper' and into Standard English. This rescues (22) above; the first article must drop, ruling out **A man bites dog*, and the second one may or may not, delivering *Man bites dog* and *Man bites a dog*.

Unfortunately this cannot be sustained. Consider for example (23):

(23) (a MAN) (BATters) (a DOG)

Here, both articles are in the 'weak' position, and so one would expect them both to obligatorily drop. However, we have seen that *Man batters a dog* is not

ungrammatical in headlines. This reformulation of Gerken does not deliver the results we require.

What about a similar reformulation of Weir (2009) — one that says that the initial syllable in an intonational phrase, if unstressed, *must* drop in headlines, while other articles may or may not? This initially seems more promising, as it seems to deliver the correct paradigm, without the problems that the reformulated Gerken analysis runs into. However, it also runs into certain problems: namely, that it mixes phonology with morphology is an unprincipled way. A metrical process does not know whether a targeted syllable is an article or not; by assumption, the only relevant information the metrical phonology can ‘see’ is whether a syllable is weak or not (and perhaps whether the syllable is part of a (grammatical) word, or stands alone). Including in this ‘explanation’ the reference to articles reduces it to question-begging, or at best to a descriptive generalisation; it is not a principled phonological explanation for the phenomenon of article drop. If weak syllables were generally permitted to optionally drop in headlines, then such an explanation might work (it would be purely phonological, and weak syllable drop has empirical support outside the issue under discussion); but this is generally not the case for headlines, although it might conceivably provide an explanation for ‘drop’ phenomena in other ‘reduced’ registers such as telegraphese and note-taking. Prepositions in headlines, for example, do not drop, although they are presumably unstressed syllables in a phonological sense; we do not see for example **PM gave bribe secretary* in the place of *PM gave bribe to secretary*.

Furthermore, if we are to take this phonological statement seriously, then we are forced to suppose that the initial syllable of *all* headlines should drop, something which is clearly not the case. Again, if we restrict the domain of the phonological process that drops the initial syllables such that it only applies to the article *a*, then we either greatly overextend the powers of metrical phonology to ‘see’ morphosyntactic information, or we reduce the statement to description, without any theoretical underpinning, simply restating that headlines cannot start with *a*.

We might here make an appeal to recoverability; articles can be recovered from context (and, in principle, this could extend to other elements dropped in headlines, such as copula and conjunctions); other elements cannot. As such, a metrical process that would prompt the deletion of ‘meaningful’, non-recoverable elements is blocked because the resulting sentence would be uninterpretable. This is a possibility, but runs the risk of being circular in defining what is ‘recoverable’ as ‘those elements that drop’. Can we really say, for example, that the preposition *to* in **PM gave bribe <to> secretary* is semantically unrecoverable or unguessable?

5.2 Phonology and string-initial ban on *a*

It does not seem as if phonology can provide an explanation for article drop in the general case. But can it provide an explanation for the ban on string-initial *a*? This initially seems much more promising. The word *a*, /ə/, is the phonologically weakest word that can be imagined in English. Could there be a phonological principle that insists that this word does not start a phonological phrase in headlines, or is deleted just in case it is string-initial?

This is very possible, and I suggest it as an avenue for future work on this issue. However, the principal problem with this suggestion is that phonological rules of this sort very rarely *enforce* deletion. They may *permit* it; and the phonological forms generated by deletion may indeed be the normal ones. However, a phonological deletion process that actively *rules out* the non-deleted variant — particularly a metrical process of dropping syllables — is at the least a strange rule to posit. (Of course, in rule-based phonology operating at the *phonemic* level, obligatory deletion is not uncommon, but I am considering cases such as the metrical explanation for subject drop; while the dropping process permits subject drop in English, it by no means requires it.)

Furthermore, it is questionable to what extent saying that ‘phonological processes result in the deletion of initial *a*’ is not simply restating the generalisation. *How* does ‘phonology’ do this? And why only to the article *a*? Why not, for example, obligatorily delete other instances of string-initial /ə/, such as for example in a possible headline *Astonishing rates of growth claimed*? Why is this not *Stonishing rates of growth claimed*, with obligatorily elided initial /ə/? There are various possible reasons why not: this could be a case of typographical convention, or the generalised process of article drop may be ‘interfering’ in some (as yet ill-defined) way to ensure that only the article *a*, and not other instances of the syllable /ə/, is obligatorily dropped. As stated above, exploring this possibility is left to future work.

6 Semantic accounts

A semantic theory of article drop would appeal to meaning, or an interplay between meaning and structural position, to provide an explanation for the distribution of dropped articles in headlines. I will consider the account of de Lange (2004), specifically relating to headlines, but ultimately reject it. I will then consider the infant-speech-based account of article drop proposed by Baauw et al. (2002), and suggest that this analysis will go some way towards solving the problem of article drop in headlines, but will not provide a full explanation.

6.1 De Lange (2004)

De Lange (2004), as discussed above in the literature review section, contends that article drop in English can be explained by reference to information theory, and argues that articles in sentence-initial position bear a low information load and can therefore be dropped.

The initial problem with this argument is that it does not account for the fact that (24 a) below is just as grammatical as (24 b), if not more so.

- (24) (a) Man bites dog
 (b) Man bites a dog

If articles outside of sentence-initial position have to bear discourse features and therefore have a relatively higher information load than articles in sentence initial position — and are therefore more resistant to drop — then why is (24 a) grammatical? What allows the second article — the one ‘belonging’ to *dog* — to be dropped? And why is that in fact (at least by this author’s intuitions) the *preferred* pattern? Should we not expect, if anything, the opposite intuition — for (24 b) to be preferred to (24 a)?

It may be argued that context allows us to supply the relevant discourse features, so the articles have little informational load anyway, and can (and should) be dropped. But, if the reader can always be relied upon to supply the discourse features, then what privileges the sentence-initial position as the position where ‘old information’ resides? What is left of de Lange’s explanation in that case?

A further argument against de Lange’s position is the one already briefly mentioned; that is, that it is just not true that the sentence-initial position is the host for old information. Or, rather: if there is old information, then it is very possible that the sentence-initial position is the canonical location for it, but the converse does not hold; we cannot assume that anything that is in sentence-initial position is old information. Indeed, in headlines of the type in (24 a), we see this to be the case. The man in question is (presumably) not a man previously known to the reader. The form with an article would be *a man*, not *the man*. De Lange’s approach relies on the reader being able to predict, merely by virtue of the

position of the sentence-initial element, that it is old information (and therefore does not require to have its discourse features encoded on an article); but the reader *cannot* predict that from (24 a), and if the reader tried, it would likely be the *wrong* prediction. Again, one might argue that context allows one to guess whether the man is known or not (here, if the man were known, presumably the headline writers would use a more specific noun than just *man*, for example). But, again, if we fall back on a contextual explanation, we lose any explanation for why the sentence-initial position is privileged.

Furthermore, de Lange's theory cannot account for the asymmetry between initial *a* (illicit) and initial *the* (licit). In fact, insofar as de Lange's theory makes a prediction, it is again completely wrong. If the fact of a DP representing old information can be predicted from initial position, then in fact *a* should be *undroppable*, because it is the only way in which the reader can interpret the sentence-initial DP as representing new information.¹⁰ For those reasons, the semantic/information-theoretic account of article drop proposed by de Lange must be rejected.

6.2 Baauw et al. (2002)

Baauw et al. (2002) study the speech of Dutch-speaking infants and agrammatics. They argue that there is a correlation between the use of articles and tense. Baauw et al. argue that both articles and tense serve to act as links between NPs and VPs, and the entities and events (respectively) that they denote. So, while the NP *man* is only semantically linked to a concept of MAN, the DP *a man* picks out a specific man; and while the VP *bite the dog* is only semantically linked to a concept of dog-biting, the tensed VP *bit the dog* picks out a specific event of dog-biting. The metaphor that Baauw et al. use, following Heim (1982), is that of a 'file card' semantics; articles and tenses act as pointers to a specific 'file card' designating an individual or event, respectively. However, there are other ways of introducing file cards, extra-syntactically, relying on discourse; Baauw et al. (2002) argue that this provides an explanation for determiner drop and non-tensed sentences in the adult language, such as the below (their (4, 5) combined):

- (25) (a) Q: Wie heeft jou gisteren gebeld?
 'Who called you yesterday?'
 A: Oh, *meisje* van school.
 Oh, girl from school
- (b) *Deur* dicht!
 door shut
 'Shut the door!'

¹⁰Thanks to Klaus Abels for pointing this out.

- (c) Leuk *huisje* heb je.
 nice house have you
 'Nice house you have.'
- (d) Ik een huis *kopen?* Nooit!
 me a house buy-INF never
- (e) Maria vertelde Peter een mop. Hij *lachen*.
 Maria told Peter a joke he laugh-INF

Baauw et al. (2002) argue that article drop and tenseless sentences in infant and agrammatic speech result, not from a lack of ability to produce the functional structure (DP and TP) corresponding to the dropped elements, but from an overgeneralisation of the ability of discourse to introduce such 'file cards', obviating the need for articles or tense. Only later in their linguistic development do infants realise that such extra-grammatical introduction of 'file cards' is restricted greatly by context, and start to consistently introduce articles and tense in all cases (except those where the adult language does not require it).

Can this help our analysis of English headlines? Obviously the extent to which it can directly apply is limited. English is not Dutch, and English does not demonstrate article drop in normal adult speech (or at least, not in the same way as Dutch does; there are certain cases in English like *Nice house you have*, but these may be better explained as a phonological dropping process à la Weir 2009). And while English does show (bare, *to*-less) infinitives in sentences like *Me, buy a house? Never!*, English does not generalise this to sentences like *Mary told Peter a joke. *He laugh*.

Nevertheless, the argument of Baauw et al. may still provide insights for an analysis of English headlines. Let us suppose that the grammar of headlines is one that very freely allows extra-syntactic introduction of 'file cards', at least to refer to individuals. We then have a grammar which allows very free drop of articles, except in 'extraordinary' situations such as that of apposition (*Sally the Streaker*), where the determiner may be argued to be playing a syntactic role independent of simply establishing the discourse status of the NP (see Matushansky (2005) for some discussion on the role of articles in such cases).

Note that such an explanation does not reduce simply to the pragmatic 'delete articles wherever', which we have seen is inadequate (it does not explain the apposition, etc. cases). Nor is it a restatement of a vague (and arguably circular) 'things that are recoverable can be dropped' rule (as we have discussed, this does not explain why we do not see e.g. preposition drop). The argument here specifically restricts itself to the dropping of articles. Neither is it a restatement of de Lange (2004)'s arguments concerning information content, where articles were dropped in order to increase the mean information content of a headline. Recall that de Lange's arguments made incorrect predictions concerning the relative

goodness of examples such as *Man bites dog* vs. *Man bites a dog*.

This could provide an explanation for why article drop (and not, say, preposition drop) happens at all in headlines. But does it provide an explanation for the restriction on the phenomenon; the restriction on string-initial *a*? Baauw et al. do note a subject-object asymmetry in Dutch child speech, but curiously, this asymmetry is the opposite way from that which we might like to see to explain the English case; in Dutch child speech, it is *objects* which have the tendency to drop articles more often, not subjects. Baauw et al. explain this by suggesting that Dutch-speaking infants overgeneralise from sentences in (adult) Dutch where count nouns may appear without determiners in object position, or after prepositions, such as *Peter reed auto* 'Peter drove a car' (Baauw et al.'s (8b)), where *auto* 'car' appears without a determiner. However, this explanation cannot help us determine why in English headlines, there is a ban on string-initial *a*.

More generally, it seems unlikely that a semantic explanation could provide an explanation for a phenomenon which makes a specific reference to the stringwise position of an element, rather than to the semantics of that element. Clearly there are correlations between the relative positioning of elements in sentences and elements of their semantics, and features such as whether they represent old or new information. But, as we have already discussed when considering de Lange (2004), such correlations do not seem to provide any satisfactory explanation of the ban on string-initial *a*. A generalisation of Baauw et al.'s arguments concerning article drop in Dutch children may successfully explain the phenomenon of article drop in English headlines at the general level; however, it does not explain the constraint on the indefinite article string-initially.

7 Syntax

So far, we have speculated that the possibility of generalised article drop in headlines may be licensed by a semantic procedure. However, we do not have an explanation for the apparent string-initial ban on the article *a*.

It is questionable to what extent a syntactic explanation, which by its nature would have as its domain the syntactic structure built ‘above’ the level of strings of tokens, could explain a phenomenon which appears to have as its domain the string itself — a generalisation stating, essentially, ‘don’t start with *a*’. However, work has been previously done on certain characteristics of the language characterised as “block language” by appealing to a *truncation* analysis, where high branches of the syntactic tree are not projected in certain ‘reduced’ registers of language.

7.1 Truncation

In particular, Rizzi (1994) proposes a truncation analysis for the case of null subjects in infant speech. Under Rizzi’s analysis, the whole of the CP layer is not projected in infant speech. According to Rizzi’s argument, the lack of any syntactic structure above the sentential subject position in a clause licenses the presence of an empty category in that subject position; Rizzi reinterprets conditions on the government of empty categories, to have them stipulate that empty categories must be governed by antecedents *if they can be*. An empty category at the extreme left edge of a clause, with no syntactic structure above it, could not be governed by any antecedent, and so does not need to be.

Notably, Haegeman (1997) extends this analysis to the case of null subjects in English and French diaries, suggesting that a similar process is at work in these written registers. In diaries, the CP layer can optionally be similarly truncated, which allows null subjects such as in the classical diary drop case. Haegeman supports this by demonstrating that subject drop in diaries is incompatible with phenomena which require the CP layer to be present, such as topicalisation or question formation.¹¹

It is tempting to try to extend a truncation analysis to the present problem of obligatory article drop; perhaps to postulate that in headlines, high levels of syntactic structure are not projected, such that the article *a*, if at the left edge of the syntactic tree, is obligatorily ‘truncated’ and as such is not seen in headlines. Below I present one attempt to extend the truncation analysis to the case of headlines. While it would be, in my opinion, a very satisfying outcome to unify

¹¹In fact, the analysis is somewhat more detailed than this, in order to accommodate the fact that adjunct topicalisation is compatible with subject drop in diaries (*Tomorrow, ∅ will go to gym*) while argument topicalisation is not (**More problems, ∅ don’t need*). The reader is referred to the Haegeman paper for full discussion.

the analyses of the cases of headlines, ‘diary drop’, and other examples of “block language” by an appeal to truncation in all cases — for reasons which I outline in section 8 of this dissertation — unfortunately the attempt presented in this dissertation fails for various reasons. However, hopefully this attempt will inspire future work on the syntax of headlines and string-initial *a*-dropping.

7.2 High determiners

The first hurdle that a ‘truncation’ analysis has to contend with is that determiners are standardly not supposed to be projected as an independent functional projection that can be truncated in the same way that, for example, the CP layer can be. Rather, determiners are the head of DP, taking their associated NP as a complement; and DPs act as one unit for syntactic purposes. It is difficult to see how a determiner could be ‘truncated’ in this framework while maintaining the overall structure of the rest of the clause.

A possible solution to the problem is simply to challenge the assumption that determiners are not functional projections; in various works, for example Sportiche (2005), Sportiche does exactly this. Sportiche posits that determiners are not generated with NPs inside the verb phrase, but are rather generated high in the clause; the NPs become associated with the determiners by movement. Specifically, the determiner of the sentential subject is generated high up in the clause, outside of the VP; while the determiner of the sentential object is generated in one of the high Larsonian shells of a VP. So, for example, the structure of the sentence *the boy opened this door* is as below (see Sportiche 2005:53):

(26) the [_{VP} [_{NP} boy] [V₁ [the [_{VP} [_{NP} door] V₂]], where V₁ + V₂ = open

This view of determiners is based on a detailed argument concerning the possibility of reconstruction of NP to a lower position in the clause, and the impossibility of reconstructing the restriction of the determiner to a low position. The details of Sportiche’s analysis of determiners will not be gone into here; the reader is referred to Sportiche (2005) for the full detail of the proposal. If, however, we decide to adopt this view of determiners, then the question of articles in headlines may be treatable using a truncation analysis.

7.3 Details of a truncation analysis

Sportiche argues for a high positioning of determiners on semantic grounds, not syntactic grounds; as such he does not specify the precise position which the determiner that ‘belongs’ to the subject would be generated in. Let us suggest, for concreteness, that the D ‘belonging’ to the sentential subject projects a functional

projection DP in the position immediately above AgrSP.¹² Let us now make a truncation proposal:

(27) *Truncated block language*

In headlines, as well as in other examples of “block language” in the sense of Straumann (1935), the syntactic tree above AgrSP¹³ is not projected; all levels of syntactic structure above AgrSP are truncated.

What are the consequences of extending such an analysis to the headlines case? The immediate one is the desired consequence: subjects are bare NPs, without determiners. The determiner that would ‘belong’ to the subject would have no site to be generated in, assuming that that site is above [Spec, AgrSP].

If headlines has a grammar containing the principles above — the generalised possibility of determiner drop (possibly due to a very free procedure of “extra-syntactic” introduction of “file cards” referring to individuals into the discourse), and the truncation of syntactic structure above the sentential subject position — then we can explain the paradigm we see below:

- (28) (a) A man bites a dog
 (b) Man bites a dog
 (c) Man bites dog
 (d) *A man bites dog

(28 a) is grammatical, but (arguably) is simply code-switching from standard written English, and isn’t an example of headlines at all. (28 b) is an example of headlines: the subject NP is bare, which is legitimate, both because the semantic function of the article is introduced “extra-syntactically”, as discussed above, and because there would not be a position for the article to be generated in even if it were required (due to the truncation of the syntactic tree above the level of the subject). (28 c) has two bare NPs, both of which are legitimate for similar reasons (although the object NP must be determiner-less because of the “extra-syntactic” introduction of the function of the article, rather than because there is not a position for the article in a D⁰ node; by assumption, such a position would be within the VP and so is not truncated).

(28 d), by contrast, is ungrammatical. If a subject has a determiner, as in (28 d), then high positions in the tree cannot have been truncated (positions above the sentential subject position must have been projected, including a location for the generation of the subject’s D). In this case, what we are dealing with is not

¹²I do not mean to make this analysis greatly theory-specific, or indeed to prejudice any element of Sportiche’s analysis; nothing hinges on the choice of AgrSP particularly. The point is to label that position in the clause that is the highest position for the NP subject before it adjoins to its determiner.

¹³The same comment applies as in the previous footnote.

headline, which obligatorily truncates such high positions in the tree. But if this sentence does not belong to headlines, then it cannot permit the “generalised” article drop either. As such, the lack of a determiner on the object renders (28 d) ungrammatical.

7.4 Problems

While the truncation analysis outlined above is tempting, it has many problems which conspire to make it unworkable. One of the most obvious is that it does not appear to make a distinction between the articles *a* and *the*, or indeed between articles and other types of determiners. There is no explanation as to why it is specifically the article *a* that obligatorily drops in headlines, not *the*, or other determiners. Under this analysis, it is the D^0 node that is truncated, but that node is the host for a far larger number of parts of speech than just articles. However, we do not seem to see generalised *determiner* drop in headlines.

A second problem is that of positioning of the determiner. Even if one accepts Sportiche’s analysis that determiners are not, in base position, sisters of their related NPs, and are rather generated as part of the functional structure of the clause, there are still limits on how high in the clause determiners are generated. In particular the contention above, that the D node is generated above the node marking tense, is not tenable. Obviously, for a truncation analysis to work, the D node must be generated above the tense node; if it were not, the truncation in the headline register would truncate the tense node as well as the D node. However, we clearly do not see this; the tense node can be present even in cases where the initial article is not present:

- (29) (a) Man [_T had] bitten dog before
 (b) Man [_T might] bite dog again¹⁴
 (c) Man [_T will] bite dog at full moon

So, the determiner must be generated above the tense node for a truncation analysis to work. However, under Sportiche’s analysis, it cannot be. One of the principal reasons Sportiche gives for a high position for determiners (see Sportiche 2005:§5) is that certain determiners cannot reconstruct to the base position of a raised subject. Consider the below example (Sportiche 2005’s (64), adapted):

- (30) (a) In 1986, no integer had been proved to falsify Fermat’s theorem.
 (b) In 1986, no *x* had been proved [integer *x* to falsify Fermat’s theorem]
 = In 1986, there had not been found an integer that falsified Fermat’s theorem.

¹⁴There is a confounding factor with both of the first two examples; namely, that these headlines impose a strong sense of definiteness on the subject — that is, the missing article is *the*, not *a*, which is the article that provides the problem. I believe this is not the case with (29 c), however.

- (c) In 1986, had been proved [no integer to falsify Fermat's theorem]
 = In 1986, it had been proved that no integer falsify Fermat's theorem.

Sportiche argues that only (30b) is a possible reading for (30a); (30c) is not a possible reading. In other words, the determiner *no* cannot reconstruct to the base position of the subject *integer*. Sportiche argues that this shows that the NP *integer* is not, in base position, the complement of the D *no* in these sentences; rather, *no* is generated high in the structure, and *integer* raises to adjoin to it.

However, if we accept this logic, then we can also demonstrate that the determiner *no* is generated *below* the tense node. We can demonstrate this by showing that the determiner *can* reconstruct to below the scope of modal verbs occurring in the tense node. Consider the below example:

- (31) (a) We don't know what the jury is going to decide yet. **No suspect might be found guilty.**
 (b) no *x* might [suspect *x* be found guilty]
 = for no suspect *x* is it possible that *x* will be found guilty
 (c) might [no suspect be found guilty]
 = it is possible that no suspect will be found guilty

In my estimation, the sentence in (31a) has (31c) as a possible reading. (My intuitions are unclear about the reading in (31b), but for the present problem this is a moot point.) If we accept Sportiche's logic concerning reconstruction, then such a reading indicates that the determiner *no* must be generated *below* the tense node containing the modal verb *might*. As such, we have a contradiction; in order for the truncation analysis to work in the headlines case, the determiner must be generated above the tense node, but to satisfy the constraints on reconstruction, the determiner must be generated below the tense node.

There is yet a further problem with a truncation analysis. Consider the below headline:

- (32) *A man bit dog, claims source

It is not the case that (32) is ungrammatical because the high functional structure 'should not' be present (because headlines is a truncated register) and therefore the determiner *a* has no place to be generated. In a quotative structure such as (32), the quoted material has undergone movement similar to topicalisation (see Collins 1997:ch. 3 for discussion of the point). If this is the case, then the truncation analysis cannot 'prune' the *a* from *a man*, as that chunk of structure is not generated as a matrix subject (indeed, the matrix subject of (32) is *source*). Again, the core point is that the restriction is not one on matrix subjects (as the truncation analysis would predict), but rather apparently on string-initial occurrences of *a*, which the truncation analysis appears ill-equipped to handle.

7.5 Possible further work

It is clear that the truncation analysis suggested above contains too many flaws to be considered as an explanation in its current form. However, some form of truncation analysis of headlines would be satisfying, for reasons I shall explore in the next section. One possible avenue for future work would be to abandon reliance on Sportiche's theory of high-generated determiners, and consider the possibility that it is in fact DPs which undergo truncation in headlines (as the whole clause is, according to Haegeman (1997), in the diary drop case), with the determiner being truncated (or, alternatively, an empty determiner position being licensed in some other way, as is the case for null subject under Rizzi (1994)'s and Haegeman (1997)'s analysis). Such an explanation would, however, have to find some independent way to explain why this would only apply to those DPs with indefinite articles in string-initial position in headlines (if, indeed, that is the correct generalisation; it may be that more careful empirical work will show that it is not).

8 Other forms of “block language”

The analysis proposed above for headlines has wider implications for the study of written language, and in particular what Straumann (1935) terms “block language”, the ‘reduced register’ of headlines, diaries, and so on. In this section I will consider what the implications of a successful truncation analysis of ‘reduced register’ would be, and why it would be desirable. In particular, I will consider the acquisition of the syntax of “block language”, and whether a truncation analysis is sufficient to explain characteristics of block language beyond the cases already described of article omission in headlines and subject omission in diaries.

8.1 Acquisition

In my previous work concerning the ‘reduced’ written register, specifically concerning the phenomenon of “diary drop”, I suggested that there were unanswered questions concerning the differences of syntax between spoken and written language (Weir 2009:29). Specifically, I noted that the set of grammatical sentences in written language (at least in written English) appears to be a superset of the set of grammatical sentences in spoken language, as there are sentences in written language which are ungrammatical in spoken language, such as sentences with omitted subject in diaries, or headlines with omitted articles. I questioned how humans are able to learn the syntactic properties of written language and indeed gain intuitions about their grammaticality, despite the fact that written competence is acquired long after the spoken language is acquired. The thesis suggested in the present work, that a chief characteristic of block language is the truncation of high layers of syntactic structure, may go some way both towards providing a unified explanation for the characteristics that appear to link different manifestations of block language, and may also explain issues of learnability and comprehensibility.

A truncation analysis of block language does not suppose that the reader of block language requires to learn any ‘new’ syntactic principles in order to understand it. It is, literally, a ‘reduced’ variant of the syntactic competence that they already possess. All that needs to be learnt is the fact that block language truncates the syntactic tree above a certain point; in all other respects, the syntax is that of English. Obviously this fact is not consciously learnt as such, but rather absorbed as an alternative syntactic configuration (in the same way that English-speaking learners of French do not consciously learn that verbs move to the T position in French; but they do learn the fact that, for example, adverbs appear after verbs in French, and they absorb the corresponding syntactic configuration into their linguistic knowledge). Given that the distinction between spoken English and the block-language register of written English would be based on such a relatively small difference in syntax, the acquisition of such a difference

(given sufficient exposure to examples of block language such as headlines) would not be hard, even though humans learn the syntax of written language after the 'critical period' of language acquisition. Arguably, the acquisition of the separate syntax of block language is comparable to the acquisition of the syntax of a second language.

8.2 Other characteristics of block language

If a truncation analysis could explain article drop in headlines, as well as diary drop as Haegeman (1997) argues, then it would be tempting to suggest that this is the defining difference between block language (at least in English) and other forms of English, and that other facts about the syntax of block language can be derived from it.

There are, however, various other properties of block language that are not obviously derivable simply from a truncation analysis; for example, copula drop, or conjunction drop. There is not space here to speculate on ways in which such phenomena could be explained by a truncation analysis. It is very possible, however, that they cannot be. In order to explain article drop in the general case in headlines, we have already had to appeal to a generalised procedure of article drop, whereby determiners are not required if the 'pointer' that they would otherwise introduce is introduced extra-syntactically through discourse. So truncation procedures, if they do indeed happen, cannot be the only factor distinguishing block language from spoken language. However, can a truncation analysis go a long way towards explaining other features of block language?

In this regard, Haegeman (1997:262ff.) speculates that a truncation analysis may explain such features as copula deletion. Haegeman notes that null copulas appear to be subject to much the same syntactic constraints (incompatibility with topicalisation or question-formation, for example) as subject drop; the same syntactic constraints that arguably act as diagnostic of a truncated structure without CP layer. Haegeman proposes that null copulas are subject to similar licensing conditions as null subjects; they are only licit if there is no possible higher head to govern them. In a truncated register, the copula occupies the highest head position; there is no head in the CP layer, as there is no CP layer. A null copula is therefore licit in much the same way as a null subject is under Haegeman's analysis. The reader is referred to Haegeman (1997) for full discussion of this point; but this should be sufficient to suggest that a unified analysis of block language may be possible, and may indeed draw heavily on a truncation analysis.

9 Conclusion

I have argued that the correct empirical generalisation to be made concerning article drop in English headlines is that articles can in general very freely drop, but that the headline must not commence with the article *a*. I have suggested that the procedure of free article drop may be explained by a mechanism proposed by Baauw et al. (2002) where the semantic contribution of the article is introduced outside the formal mechanisms of the grammar through discourse. For the constraint on string-initial *a* I have no ready explanation. I have suggested why a truncation analysis would be satisfying, but that at least one possible truncation analysis — while initially promising — does not seem to capture the full breadth of the data involved. Hopefully future work on this topic will provide an explanation, beyond a simple descriptive statement, for the ban on string-initial *a* in headlines.

References

- Baauw, Sergio, Esterella de Roo, and Sergey Avrutin. 2002. Determiner omission in language acquisition and language impairment: Syntactic and discourse factors. In *Proceedings of the Boston University Conference on Language Development (BUCLD) 26*, ed. Barbora Skarabela, Sarah Fish, and Anna H.-J. Do. Somerville, MA: Cascadilla Press.
- Collins, Chris. 1997. *Local economy*. Cambridge, MA: MIT Press.
- Gerken, LouAnn. 1991. The metrical basis for children's subjectless sentences. *Journal of Memory and Language* 30:431–51.
- Haegeman, Liliane. 1997. Register variation, truncation, and subject omission in English and in French. *English Language and Linguistics* 1:233–70.
- Heim, Irene. 1982. The semantics of definite and indefinite Noun Phrases. Doctoral Dissertation, University of Massachusetts, Amherst.
- Janda, Richard D. 1985. Note-taking English as a simplified register. *Discourse Processes* 8:437–454.
- Kostić, Aleksandar. 2004. The effects of the amount of information on processing of inflected morphology. Ms, University of Belgrade.
- de Lange, Joke. 2004. Article omission in child speech and headlines. In *OTS Yearbook 2004*, ed. Annemarie Kerkhoff, Joke de Lange, and Oren Sadeh Leicht, 109–119. Utrecht Institute of Linguistics.
- Mårdh, Ingrid. 1980. *Headlines: On the grammar of English front page headlines*. Malmö: CWK Gleerup.
- Matushansky, Ora. 2005. Iraqi head seeks arms: Are bare nouns created equal? Presentation at A Bare Workshop, December 22, 2005, Utrecht.
- McGregor, Karla K., and Laurence B. Leonard. 1994. Subject pronoun and article omissions in the speech of children with Specific Language Impairment: A phonological interpretation. *Journal of Speech and Hearing Research* 37:171–81.
- Rizzi, Luigi. 1994. Early null subjects and root null subjects. In *Language acquisition studies in generative grammar*, ed. T. Hoekstra and B. Schwarz, 151–77. Amsterdam: John Benjamins.
- Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27:379–423.
- Simon-Vandenberghe, Anne-Marie. 1981. *The grammar of headlines in The Times 1870–1970*. Brussels: Paleis der Academiën.

- Sportiche, Dominique. 2005. Division of labor between Merge and Move: Strict locality of selection and apparent reconstruction paradoxes. <http://ling.auf.net/lingBuzz/000163>.
- Stowell, Tim. 1999. Words lost and found in Headlines: The hidden structure of abbreviated English in headlines, instructions and diaries. Paper presented at York University, Toronto.
- Straumann, Heinrich. 1935. *Newspaper headlines: A study in linguistic method*. London: George Allen & Unwin Ltd.
- Waller, Brad. 1997. Against a metrical basis for subject-drop in child language. Ms, Massachusetts Institute of Technology. Available online at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.52.5111&rep=rep1&type=pdf>.
- Weir, Andrew. 2009. Subject pronoun drop in informal English. Richard M. Hogg Prize winning essay. Available online at <http://www.isle-linguistics.org/resources/weir2009.pdf>.