Current Approaches to Punctuation in Computational Linguistics

B. Say and V. Akman
Dept. of Computer Engineering and Information Science,
Bilkent University,
06533 Bilkent, Ankara, Turkey
{say,akman}@cs.bilkent.edu.tr

Abstract. Some recent studies in computational linguistics have aimed to take advantage of various cues presented by punctuation marks. This short survey is intended to summarise these research efforts and additionally, to outline a current perspective for the usage and functions of punctuation marks. We conclude by presenting an information-based framework for punctuation, influenced by treatments of several related phenomena in computational linguistics.

Key words: discourse, information, natural language syntax, natural language semantics, punctuation

Abbreviations: DRT – discourse representation theory; DRS – discourse representation structure; NLP – natural language processing; NLG – natural language generation; RST – rhetorical structure theory; SDRT – segmented discourse representation theory; SDRS – segmented discourse representation structure

This paper has not been submitted elsewhere in identical or similar form, nor will it be during the first three months after its submission to *Computers and the Humanities* without permission of Kluwer Academic Publishers.

1. Introduction

Punctuation marks have not been studied much by linguists apart from a prescriptive standpoint until the eighties. Similarly, most natural language processing systems did not take punctuation marks into account except for the period and the spacing. However, there have been recent works in linguistics (computational, corpus, and applied), giving a descriptive treatment of the role of punctuation in contemporary written language. Furthermore, various natural language processing systems have started to make use of syntactic cues provided by punctuation marks. In this short and by no means exhaustive survey, we intend to present the current state of incorporation of punctuation marks into natural language processing (NLP) systems as well as summarising the recent research (computational or within general linguistics) on descriptive characterisations of punctuation.[1] In this survey, we take punctuation marks to be not only the standard marks such as comma, colon, period, dash, etc. but also the more graphical devices such as paragraphs, tables, lists, features for emphasizing (such as use of italics).

The rest of the survey is organised as follows. Section 2 gives a current perspective on the history of punctuation and its place in writing today. In Section 3, we present some of the current linguistic studies, excluding the computational ones. In Section 4, relevant NLP work mostly on the relationship of syntax and punctuation in the area of computational linguistics is summarised and evaluated. In Section 5, semantic, intonational and discourse-wise implications of punctuation are discussed. Section 6 concludes with an information-based perspective for punctuation.

2. Punctuation and Written Language

According to Parkes (1993), the development of punctuation took place in several stages paired up with the development of the written medium. Each stage's reader group required different demands to be satisfied, thus affecting the marks and their functions. In Classical Latin writing, education was directed at preparing students for effective public speaking (Parkes, 1993, p. 5). Authors dictated their writing to the scribes and only for teaching purposes did an author, a scribe, or a corrector put different marks on the manuscript for indicating different length of pauses. Spaces between lexical words did not become customary until the tenth century (Levinson, 1985, p. 23). As opposed to punctuating for oral readers, some grammarians saw writing as a means for silently conveying meaning to the reader (Parkes, 1993, p. 21). During the eighth century, the Irish devised new graphic conventions in the written text (because Latin was mainly a written or visible language for them) and later passed those conventions onto the Anglo-Saxons (Parkes, 1993, p. 23). From 12th century onwards, a general inventory of punctuation marks was designed but, since even two scribes copying the same manuscript employed different marks, there was no standardisation (Parkes, 1993, p. 69).

When writing went beyond the boundaries of the monasteries and the clergy, and began to be used for secular purposes, economy and speed in reading became more important (Levinson, 1985, p. 38). Writers started to use punctuation to bring out the relationships between the grammatical constituents of the sentence. In particular, during 14th to 16th centuries, the humanists wanted their texts to be persuasive and demonstrative. Thus, they adopted a larger set of punctuation marks to disambiguate the logical structure of sentences. New marks corresponding to today's parentheses, semicolon and exclamation mark were devised in the 15th century. From 16th century onwards, with the widespread usage of printing a gradual standardisation emerged. Types and fonts were precut and sold to printers so the available repertory of marks was no longer personalised by the scribes. Also, before printing, the destination of the manuscript being prepared (e.g., a specific monastery or library) was mostly known beforehand. After printing became the norm, this pre-existing connection between the "publisher" and the client was broken; there was now a greater pressure for general understandability and readability of the text. The orthographic sentence became the fundamental information unit presented to the reader in an easy-to-understand manner (Levinson, 1985, p. 157). Symbols such as rhetorical question marks (adding a rhetorical effect to a positive statement), apostrophes, quotation marks, and use of italics to create emphasis emerged after the 16th century. Punctuation for the rhetorical and logical structure of the text became so widespread that 19th and early 20th century novelists frequently used punctuation as one of the features of the written medium to create illusions and "stream of consciousness" effects (Parkes, 1993, p. 87). As can be seen in the works emphasising the usage of punctuation marks in modern texts (Bayraktar, 1996; Jones, 1995; Meyer, 1986), punctuation is still an integral part of the written language. A typical English sentence is likely to contain three or four punctuation symbols, and a punctuation mark of some variety is likely to be encountered on average every fourth to seventh word (Jones, 1997, p. 86).

Thus, it is important to view punctuation from a linguistic, even from a semiotical point of view such as in (Harris, 1995). Harris does not take a writing system as being simply projected from speech. Rather, in his work written signs are analyzed according to the types of activity (forming, processing, and interpretation) they are involved in (Harris, 1995, p. 60). Writing uses spatial relations and thus is different from speech. In understanding forms of punctuation such as tabular writing, which has no counterpart in spoken language, the internal syntagmatics ("the disposition of written forms relative to each other within the graphic space" (Harris, 1995, p. 121))becomes important. For the text on the electronic media, the graphic space is dynamic (as opposed to static) and the hierarchy of the written sign is leveled out. It is the nuances supplied by punctuation that help restore that hierarchy to a degree.

3. Linguistic Perspectives on Punctuation

Style guides and grammar books (Ehrlich, 1992; McDermott, 1990; Partridge, 1953) in general give a prescriptive account of punctuation. In the applied linguistic arena there are mostly works relating to learnability. Scholes and Willis (1990) recite an experiment where university students, when asked to read a text aloud, interpreted punctuation marks as elocutionary even when the marks had other meaning changing effects. Smith (1986) describes another experiment to determine whether a graphical instruction environment is better liked by students learning punctuation. A recent ongoing project is to see how young children understand the nature and use of English punctuation so that effective ways of teaching punctuation are found (Hall and Robinson, 1996). These works provide some valuable insights as to how punctuation is learnt and perceived but to be of use from a computational point of view, more descriptive studies are needed.

The first up-to-date descriptive treatment of punctuation as a system is Meyer's PhD thesis, later published as a book (Meyer, 1983; Meyer, 1987). He concentrates on American usage of structural punctuation marks. By *structural* he means those marks that act on units not larger than the orthographic (written) sentence (thus no paragraphs) and not smaller than the word (thus no hyphens or apostrophes) (Meyer, 1986, p. 80):

> This study focused exclusively on "structural punctuation": periods, question marks, exclamation marks, commas, dashes, semicolons, colons, and parentheses. It did not deal with paragraph indentations (or separation) or apostrophes and hyphens, nor did it focus on brackets, ellipsis dots, quotation marks, and underlining, or the use of commas and colons in dates, times, etc. These are marks of punctuation whose uses have been fairly rigidly conventionalised by style manuals.

We do think that structural marks are a good working category to distinguish from text punctuation (such as paragraphs, font changes, lists) but the definition given is not exactly correct as parentheses do work on units larger than sentences. This is one of the reasons for Dale's (1991a) call for a theory of discourse (and discourse uses of punctuation) spanning the sentence boundary.

Meyer uses 12 samples, approximately 2000 words each, from the Brown Corpus (Francis and Kučera, 1982) in fiction, journalistic, and learned styles. Working on these samples he classifies and exemplifies the functions of punctuation, and how those functions are realised. Distinguishing between the functions of marks and their realisations is one important point he stresses to be usually missing from the prescriptive work. Functions basically help the reader understand efficiently and easily, emphasise a construction, or vary the rhythm of the text. He groups their realisation into two categories: marks that separate (such as periods, colons) and marks that enclose (such as dashes, parentheses). He then gives a detailed account of boundaries that punctuation marks work on: syntactic (clauses, phrases, or words), prosodic (pauses, tone units, and changes in stress and pitch), and semantic (questions, modifiers, etc.). He notes that punctuation usually overdetermines (determines more than one kind of boundary) but that it also usually favours one more than the other.

Meyer's work, to our knowledge, is the first of its kind in trying to synthesize a linguistic account of punctuation from corpus data. His book is also valuable in comparing what different style manuals prescribe and what actually happens. However, the size of the sample corpus he considers is too small (compared to what is electronically available nowadays) and the content is specifically American English. The way the linguistic analysis is presented is complete on general terms but down to specifics, it amounts to observations rather than generalizable formalisations.

Levinson's PhD thesis (1985) essentially offers a historical perspective on the development of punctuation marks. She sees two serious flaws in recent works on punctuation. One is the idea that "Punctuation marks syntax". The other is the idea that "The fundamental entity which determines punctuation is the sentence". She observes a potential circularity in that in trying to establish rules according to the distribution of punctuation, the rules require a prior notion of sentence; yet a clear definition of sentence is based on punctuation marks, namely capital letters and the period.[2] She proposes to free oneself from this circularity by separating the grammatical sentence from the orthographic one. She claims that relating punctuation to syntax may stem from the fact that it is easier to do so. Relating it with other linguistic features such as intonation contours or semantic concepts would be more difficult. She proposes to view the orthographic sentence as an "informational grouping" based on (but distinct from) syntactic structure and specified by rules of punctuation (not grammar). She defines informational grouping as putting, within the limits of the orthographic sentence, the linguistic units in the right order according to their informational links. She goes on to describe the linguistic units she uses for this purpose (i.e., proper clause structures and sentence partials) and gives a classification of the actual grouping. Sentence partials like adverbial clauses and tenseless verb phrases, as Levinson sees them, do not classify as proper clauses. In attaching sentence partials to proper clauses and to other sentence partials, a signal of attachment (an informational link) is required. Various devices can act as such a signal, viz. conjunctions, phrase ordering. Punctuation is also one of them. Consider the following examples, taken from (Levinson, 1985, p. 130), with different kinds of attachments:

(1) a. He was happy to find his book.

b. He was happy because he found his book.

c. He was happy. He found his book.

In (1c), a limit to the informational group "He was happy" has to be put by means of punctuation. Where and how a sentence partial is attached or presented gives different consequences as to different information packagings (Levinson, 1985, p. 134). Likewise, as in (2), taken from (Levinson, 1985, p. 136), sentence partials that are not felicitous on the basis of grammaticality are acceptable as instances of informational grouping:

(2) But it is a game that should be enjoyed, not taken seriously. Any more than one takes seriously the idea that the United States of Europe is just round the corner.

Levinson's work is mostly a historical account of punctuation, drawing observations on the current usage as well. Its characterizing features are distinguishing the orthographical and the grammatical sentence, and observing how punctuation facilitates information packaging. The account of sentence partials and their linking is quite vague and can be accounted for within (rather than outside, as she suggests) a grammatical formalism.

The book on which majority of the studies reviewed in the next section are based is (Nunberg, 1990). He attributes the negligence of punctuation in the linguistic community to its being relatively new as well as its being perceived as prescriptive and as a reflection of intonation. He explains that the origin of punctuation was the transcription of intonation but then the two diverged and now punctuation is a linguistic system in its own right. He describes a *text-grammar* as the collection of rules that explains distribution of explicitly-marked categories such as paragraph, sentence, or parentheticals. He intentionally excludes semantic or pragmatic relations of coherence and the like from his definition of text-grammar, as these depend on context.

Nunberg constructs his text-grammar so that it accounts for punctuation marks between text-categories (text-clauses, text-adjuncts, or text-phrases) which are themselves dealt with by the lexical grammar. He proposes various rules for English to handle the interactions between various marks. One such rule, for example, is the point absorption rule, which among other things dictates that a period will absorb a comma when they are immediately adjacent. He also touches upon the pragmatic functions of text-categories such as those separated by semi-colons. He observes that a semi-colon links two clauses in a special way (e.g., in an elaborative or, contrastive way), but the exact relation can only be inferred from context. Consider the following example which conveys both senses:[3]

(3) We preferred the mountain route to Istanbul; the highway was too crowded.

Another group of hierarchically-ordered rules are for the presentation of punctuation marks, i.e., text-category indicators — including font- and face-alterations — grouped according to whether they are linearly presented or mapped into a 2-D page layout.

Two reviews of Nunberg's book (Humphreys, 1993; Sampson, 1992) acknowledge his work rather positively. Sampson observes several counter-examples to Nunberg's rules though, drawing attention to the fact that they are not adequately based on empirical data. Switching between single and double quotations is not uniformly distinguished between American and British practices. Brackets or colon-expansions can be nested as opposed to Nunberg's suggestion (a point also noted by Jones (1997)). These kinds of stylistic choice clearly make the process of establishing a set of tidy, empirical rules for punctuation harder.

Nunberg's way (although prone to prescriptivism at times) of deciphering punctuation as a linguistic subsystem separate but related with (lexical) grammar has been a starting point for other research work to be mentioned in this survey (see also (Jones, 1996a)). When a unified theory of punctuation is born, it may not be like what Nunberg has suggested in particulars but it has to account for the issues raised by him.

In all, we see that most of these works recognise the information-providing function of punctuation marks. However, they do not attempt to provide a formal perspective (apart from Nunberg's work, which mainly covers the syntactic aspects).

4. Computational Work on Punctuation

Garside and his colleagues (1987) describe a research programme undertaken during 1976-1986 whose aim was to base NLP on the probabilistic analysis of a large corpus. In describing the tagging subsystem, they show how they use punctuation marks (tagged to delimit ambiguity). They also describe a related project on "automatic intonation assignment", which aims to produce a prosodic transcription from written forms of punctuated, spoken texts. This is, notably, one of the first studies considering punctuation in an NLP context. Another early system is the Bravice English-to-Japanese machine translation system that was developed between 1987-1991 (Fornell, 1996). This system treated punctuation marks as lexical categories integrated in the grammar. Punctuation was used to prune useless parses so that the tight memory constraints of the system (it was designed for the PCs of the time) could be overcome.

Also worth mentioning is the SUSANNE analytic scheme (Sampson, 1995). This is a notation for indicating the structural (grammatical) properties of samples of English taken from the Brown corpus (Francis and Kučera, 1982). It aims to develop a comprehensive, explicit, consistent, and theory-neutral notation that will be of use to researchers working on corpora. As part of the notation, punctuation marks have their own tags and act as leaf nodes in a SUSANNE parse tree. Various ambiguities as to where to attach them within the parse tree are worked out.[4]

Jones (1995; 1997) has done a computational analysis of the structural punctuation marks on various corpora, including the Guardian newspaper (12 million words), the Leverhulme corpus (a corpus of student essays, 356,000 words), the Wall Street Journal Corpus (184,000 words), and articles extracted from the Usenet. He made a percentage analysis of various punctuation marks used as well as comparing complexity and genre of the texts with the frequency of the marks. Bayraktar (1996) has conducted another computational study on the usage of comma, taking classes (categories of use) as specified by Ehrlich (1992) as a basis for matching 241 syntax-patterns of comma uses in the Wall Street Journal Corpus (ACL/DCI, 1991) against them. By devising a metric called *stability*, he shows how the standard usage of a semantic class can be assessed. A similar study has been attempted, again for comma categories, using the SUSANNE corpus (Ince, 1996).

There are several recent works with explicit emphasis on using punctuation in NLP. Srinivasan (1991) is interested in using punctuation for lexicography and abstracting. He stresses the need for the extraction of visual information including punctuation from texts. He further divides the functions of punctuation into four groups: delimiting, distinguishing (specifying emphasis, etc.), separating (indicating syntactic units), and morphological. His experimental work involves building of an extended lexicon for machine translation, including information extracted from the use of punctuation marks. There have also been recent research on certain punctuation marks in English. Doran's work concentrates on the role of punctuation marks in quoted speech within a lexicalised tree-adjoined grammar (Doran, 1996). Douglas and Hurst's work characterises layout-oriented punctuation devices such as tables and lists (Douglas and Hurst, 1996). A detailed analysis of the role of commas in various types of coordinated compounds is given in (Min, 1996).

A recent natural language understanding system that takes punctuation into account is the Constraint Grammar developed by Karlsson and his colleagues (1994). Constraint grammar is an effort for morphological and syntactic parsing of language-independent, unrestricted text. Karlsson *et al.* combine a grammar-based approach with optional heuristics, when the former fails. The emphasis is on discarding improper alternatives by means of constraints, which are rules for disambiguation. One of the goals of their framework is simplification of parsing through the use of typographical features such as punctuation, case (of letters), and mark-up (of texts). They treat all sentence delimiters plus non-letter and non-digit characters as specially-marked, individual words which may have features and referred to by constraints. In this way, punctuation marks are used to detect clause boundaries or lists of similar categories. Also, in recognising subjects, punctuation marks such as dashes to the left of a finite verb dramatically decrease the probability of the preceding word to be a subject. In their corpus studies, of all the finite verbs preceded by a punctuation mark, less than 5% have been found to have the preceding word as the subject. One rare example where this occurs is as follows (Karlsson *et al.*, 1994, p. 328):

(4) The company's chairman and chief executive officer, T. Marshall Hahn Jr., said the plan "isn't being adopted in response to any effort to acquire control" of the forest-products concern.

In (4) the quoted speech starts unexpectedly after the subject of the predicate.

Jones (1994a; 1994b; 1996b; 1997) describes computational parsing-related work based mainly on Nunberg's framework, using a feature-based tag grammar. He refrains from using a two-level grammar as advocated by Nunberg on the grounds that interactions between the levels make the grammar unnecessarily complex (Jones, 1994b). For Nunberg, the lexical expressions must have information about their neighbouring syntactic categories so that the text grammar can draw proper conclusions. Jones instead modifies an existing grammar for English by introducing a notion called *stoppedness* for a category that describes the punctuation mark (if any) following it. The rules cater for optionality of certain marks and the absorption rules (e.g., a period absorbing an adjacent comma) through *stop* values. Testing his grammar on the Spoken English Corpus (Taylor and Knowles, 1988), which includes varied-length sentences with rich punctuation, he concludes that the number of parses are reduced by an order of magnitude of two

(10^2) for complex sentences when using a grammar that takes care of punctuation marks as opposed to a grammar that does not. He also introduces a measure of complexity of a sentence in terms of punctuation; there is a direct relationship between number of parses a given sentence has and the average number of words residing between two punctuation marks in it. Jones revises his implementation methodology in later works (Jones, 1996b; Jones, 1996c; Jones, 1997). For instance, discarding stoppedness ensures better modularity (see presently for a discussion of (Briscoe and Carroll, 1995)). He draws 79 generalized punctuation rules from nine corpora (on colon, semicolon, dash, comma and period). The corpora based experiments lead him to hypothesize his generalizations within the X-bar theory (Jackendoff, 1977) as follows: attachment of punctuation to the non-head daughter only seems to be legal when mother and head-daughter categories are of the same level (Jones, 1996c, p. 365). His revised grammar produces similar (or even slightly better)[5] results compared to Briscoe and Carroll (1995) (see the next paragraph). Jones also gives a schematic theory of punctuation in which he classifies syntactic, semantic, and pragmatic uses of punctuation. In the latter two categories, he states that there is very little use for a comprehensive theory (Jones, 1997).

There is similar work done by Briscoe and Carroll (1994; 1995). They build a text-grammar as intended by Nunberg, by tokenising punctuation marks separately from words, and use a unification-based grammar in conjunction with a probabilistic LR parser for certain lexicographical applications in mind. Punctuation is seen as useful for not only breaking the text into suitable units for parsing but also resolving structural ambiguity. They build Definite Clause Grammar (Pereira and Warren, 1980) format rules for capturing text-sentential constraints described by Nunberg. They then integrate this grammar into another one for part-of-speech analysis. Treating text categories and syntactic categories as overlapping, and dealing with disjoint sets of features in each grammar render the integration to be more modular than the approach taken by Jones. They test the resulting grammar on the Spoken English Corpus and SUSANNE corpus and give detailed interpretations of their results according to various performance factors (Briscoe and Carroll, 1995). When about 2500 of in-coverage (covered by the resulting grammar) SUSANNE sentences were stripped off of their punctuation, around 8% of them failed to receive an analysis at all and an average sentence received 38% more parses than before. They mention further possible work to develop semantic rules for text-unit and text-adjunct combinations that have a discourse relationship by incorporating discourse relations and its interpretations.

Lee syntactically and semantically extends the grammar described above (Lee, 1995). For the semantics, she implements the distinguishing semantics between subordinating and coordinating constructs. Upon testing her grammar on a small test corpus, she finds that syntactically all the punctuated sentences have at least one parse whereas 50% of the same sentences unpunctuated do not parse at all (Lee, 1995; Briscoe, 1996).

Shiuan and Ann (1996) report an experiment about separating complex sentences with respect to punctuation and other *link words* and parsing the so-created chunks first. They report a 21% error reduction in parsing as compared to the performance of their original, non-divide-and-conquer parser. Osborne (1996) recites an experiment where even a simplified model of punctuation enhanced learning unification-based grammars. White's work (1995) examines punctuation from a Natural Language Generation (NLG) point of view. He investigates how Nunberg's approach to presenting punctuation (and other formatting devices) might be incorporated into NLG systems. He extends and criticises Nunberg's analysis of punctuation presentation rules, giving examples where some options work fine from a parsing point of view but overgenerate from a generation point of view. He then proposes a layered architecture for implementation. His architecture has three components: syntactic, morphological, and graphical. These deal with punctuation presentation rules for hierarchy, adjacency, and graphical form, respectively. In this way, White aims to put rules in the process of generating punctuation into action as early as possible, thus overcoming some of the shortcomings of Nunberg's framework.

As can be seen, there is considerable recent work on using punctuation marks especially for the task of syntactic parsing and characterising their usage with computerised corpora. As to the systems described (Garside *et al.*, 1987; Karlsson *et al.*, 1994), it is hard to evaluate to what degree they incorporate punctuation without actually working on them, but one can at least say that they have taken some steps towards such an incorporation. From a parsing point of view, Briscoe and Carroll's (1995) and Jones' (1997) systems are significant and comparable. More work on specific marks such as quotations (Doran, 1996) will still be valuable. The next question may be whether the works cited above cover enough ground to characterize punctuation.

5. About Other Aspects of Punctuation

Consider the following sentences from Nunberg (1990, p. 13):

(5) a. Order your furniture on Monday, take it home on Tuesday.

b. Order your furniture on Monday; take it home on Tuesday.

Nunberg indicates that (5a) has a conditional sense whereas (5b) is merely a conjunction of the two sentences. Now consider the following sentences again from Nunberg (1990, p. 31):

(6) a. He reported the decision: we were forbidden to speak with the chairman directly.

b. He reported the decision; we were forbidden to speak with the chairman directly.

In (6a) the spokesman announced the decision and the decision was that they were forbidden to speak with the chairman directly. In (6b) the spokesman reported the decision to the chairman as others were forbidden to speak with the chairman directly. In a less intuitive setting, (6b) can also mean that the reason the spokesman announced the decision himself (rather than the chairman) was that they were forbidden to speak with the chairman directly.

In the made-up example below, dashes indicate an abrupt change of subject, which acts as a discourse segment by itself:

(7) Now, I shall tell you the full story — but first have another cup of tea.

These examples suggest a relationship between discourse and punctuation. Dale in (1991a; 1991b) raises questions about what roles punctuation plays within discourse structure. He points out the relationship among lexical markers[6], punctuation marks, and graphical markers (such as paragraph breaks or lists) within the structure of written text. Punctuation marks are not openly linguistic as cue words nor openly layout oriented such as lists but they at times perform similar functions. He observes that many uses of certain marks (colon, semicolon, dash, parentheses, comma) act as signals of discourse structure usually within the orthographic sentence level. This justifies the need of a discourse theory that should be able to operate below and above the orthographic sentence level. Particularly, Dale states that punctuation underdetermines rhetorical relations in a text since the same marks can be used for different relations (as noted by Nunberg). This urges Dale to consider the possibility of taking a syntactic view of punctuation within discourse. This might involve, for example, determining whether one segment serves as a precondition for another without assigning exact coherence relations.[7] He tries preliminaries of both an intentional structure and a coherence structure by respectively using the Theory of Discourse Structures (Grosz and Sidner, 1986), and the Rhetorical Structure Theory (RST) (Mann and Thompson, 1987). RST involves characterising coherence relations that hold between arbitrarily long units of text. Relationships are numerous (including elaboration, justification, etc.) and can be applied hierarchically. RST is particularly useful in written discourse analysis and has also been used for NLG systems. Dale tentatively classifies punctuation marks into three main groups: juxtaposed elements that provide rhetorical balance, those that indicate strength of the relation, and those that carry special rhetorical relations such as elaboration.

Pascual and Virbel (1996) analyse the "textual" punctuation marks (such as paragraphing, indentation, and font changes) in text understanding and generation, from a semantic point of view. They call certain entities (such as chapters, introductions, theorems) *textual objects* and define a textual architecture by means of metasentences that describe the positional, typographical, and speech-act based relations between those objects distinguished by textual punctuation marks.

There is also a parallel between intonation (and the attempts to formalise it) and punctuation. Cruttenden (1986) explains that for many uses of punctuation there is no intonational equivalent. Some exceptional uses usually correlate with the boundaries of a separate intonation group such as a pair of commas in parenthetical use. He claims that the often unnecessary usage of a comma between the subject and the predicate of a clause occurs from such a coincidence. Bolinger (1989), on the other hand, has investigated the relationships of intonation to discourse and grammar. He finds that intonation and grammar are pragmatically (but not linguistically) interdependent, but this interdependence is not a strict one. He gives examples where punctuation marks help clarify the intonation, but in written text intonational information is bound to be lost even with punctuation. "I told the doctor I was sick!" would certainly be read with a different intonation if one knows it is written on a tombstone (Bolinger, 1989, p. 68). Chafe (1988) has done experiments in explicating the relationship between punctuation and intonation. He claims that there is a "covert prosody" of written language which affects both the writers' and the readers' imagery, and some of it is made explicit by punctuation. His experiments include reading aloud and inserting punctuation in text from which the original punctuation has been removed. He concludes that "punctuation units" (stretches of language between punctuation marks) can be considerably longer than "intonation units" of speech due to the nature of writing. Every punctuation unit boundary is not found to be an intonation unit boundary either as there are other functions of punctuation marks. Intonation is also a device that can bring informational cues to the analysis of spoken speech. This has been explored by Oehrle (1995) and Steedman (1991) using Combinatorial Categorial Grammar to capture the syntactic and intonational structures of English. Prevost and Steedman (1994) use such a combined framework of prosody and information structure (theme-rheme) to assign natural and correct intonational contours within the generator of a text-to-speech system to responses to prosodically annotated database queries. These works are also related with Vallduví's work (1992) in bringing out the information-based structure of the sentence.

Punctuation's significance does not fall straight into one compartment of linguistics. This is the main contribution of the works cited in this section. Punctuation marks are used to make the text maximally relevant and informational for the reader. To formalize punctuation marks from an information-based perspective will thus be beneficial for a unifying framework.

6. An Information-Based Perspective

Punctuation marks can be seen as contributing to the information conveyed by the sentence (or intrasentential clauses) to the discourse. Vallduví and his colleagues' (Vallduví, 1992; Engdahl and Vallduví, 1996) treatment of *information packaging* may be adjusted to include the effects of punctuation marks. By information packaging he means the non-truth conditional meaning of a sentence and how it is brought about. Information is defined as the propositional content which constitutes a contribution of knowledge to reader's knowledge store. Vallduví gives the following examples (Vallduví, 1992, p. 2):

(8) a. He hates broccoli.
    b. Broccoli he hates.

(8a) and (8b) are truth-conditionally equivalent but they say what they say about the world in different ways. Vallduví devises a scheme of focus-ground that accounts for these differences in information packaging. How information packaging is realised linguistically (i.e., by means of intonation, syntax, morphology, or punctuation) may differ from language to language.[8] As Levinson (1985) indicates, it is important to see punctuation as a device operating on the written sentence as opposed to a grammatical sentence, sorting out the interrelationships between the informational groups.

The ideas on the various phenomena described above can be integrated for punctuation from an informational perspective. We (Say, 1995; Say and Akman, 1996a; Say and Akman, 1996b) have attempted to draw preliminaries to such an approach by using variants of Discourse Representation Theory (DRT) (Asher, 1993; Kamp and Reyle, 1993). We basically model the use of punctuation marks that change the structure of the discourse within or above orthographic sentence. DRT has been designed to provide a systematic and adequate account of the truth conditions of multi-sentential discourses (Kamp and Reyle, 1993). In DRT, representational boxes called Discourse Representation Structures (DRSs) are built up while the discourse is being interpreted. A DRS has a set of discourse referents (entities or eventualities in the discourse) and a set of conditions that state certain properties and relations relating to those discourse referents (which can be DRSs themselves). A DRS is built incrementally to cover the whole discourse with guidance from a DRS construction algorithm on the basis of the syntactic structure of a sentence. DRSs can be translated into first-order logic with classical model-theoretic semantics. Accessibility conditions determine the scoping of referents for possible anaphora resolution.

DRT has not only well-developed accounts for anaphora, quantification, tense, etc., but also applicability in a strong computational sense. However, it lacks, in its bare bones version (Kamp and Reyle, 1993), constructs that deal with the structure and the relations of the discourse, which are required for certain usages of punctuation. Such constructs are provided by Asher (1993) within a related theory he presents for discourse structure for analyzing abstract entity anaphora. The structure and the segmentation of discourse may help to choose antecedents for anaphoric reference. The basic entities at this level are called *segmented DRSs* (SDRSs) by Asher. They are imposed on the logical structure created by DRSs by relating DRSs with discourse relations, which act as conditions for SDRSs. Built incrementally as DRSs, a unit of information is defined to be a *constituent*. As opposed to core DRT, SDRT has one sentence as a constituent and builds a structure for discourse compositionally from there. For our purposes, the basic constituent of a SDRS does not have to be a sentence but can be a text-clause or a text-phrase separated or enclosed by punctuation marks. Asher uses a subset of relations essentially borrowed from RST (Mann and Thompson, 1987) to link SDRSs. Accessibility in SDRT is rather like the accessibility conditions of merged DRSs in Compositional DRT (Muskens, 1996) in that a grand union of all the constituents and their constituency relationships determine what antecedents are available for a particular referent. In addition, however, there is a principle of availability determined by the discourse structure: "A discourse referent may find an antecedent either in a constituent that is connected by a discourse relation to the current constituent or in a constituent that is the topic of the current constituent" (Asher, 1993, p. 314).

x y z X Z W	x y z X Z W
Jane(x) Joe(y) Sue(z) $X=y \oplus z$ $Z=x \oplus X$ books-on-England(W) write(Z,W) [U | x's books(U), $U \subseteq W$, be(U,best-sellers)] $\Rightarrow$ [be-jealous(X)]	Jane(x) Joe(y) Sue(z) $X=x \oplus y$ $Z=X \oplus z$ books-on-England(W) write(Z,W) [U | z's books(U), $U \subseteq W$, be(U,best-sellers)] $\Rightarrow$ [be-jealous(X)]

Figure 1. DRSs for (9a) and (9b)

Considered below are several types of punctuated sentences that influence the semantics and the pragmatics of the discourse. We do not intend to automatically extract SDRSs from punctuated sentences. At best, we can extract templates that might work in sync with such a module of a system. Even as such, we hope to get a twofold benefit by using the SDRT framework: first, by adequately specifying the semantic and discourse related functions of punctuation, and second, by suggesting some revisions to how SDRT or DRT deals with punctuated written text as we examine corpora. We briefly comment on the punctuated sentences below to show how they can be dealt with DRSs or SDRSs. (To avoid cluttering, tense and various other information have in general been omitted from the following DRSs.)

(9) a. Jane, and Joe and Sue write books on England. If her books are best-sellers then they are jealous.

b. Jane and Joe, and Sue write books on England. If her books are best-sellers then they are jealous.

The exact position of the comma in the first sentence changes the resolution of pronominal anaphora in the second sentence, which is identical in both pieces of discourse. (9a) will have *her* attached to Jane and *they* to Joe and Sue, whereas (9b) will have *her* attached to Sue and *they* to Jane and Joe. This can also be dealt with plain DRSs as shown in Figure 1.[9]

(10) John — his brother is also an athlete — won the university medal easily. He is an ambitious guy.

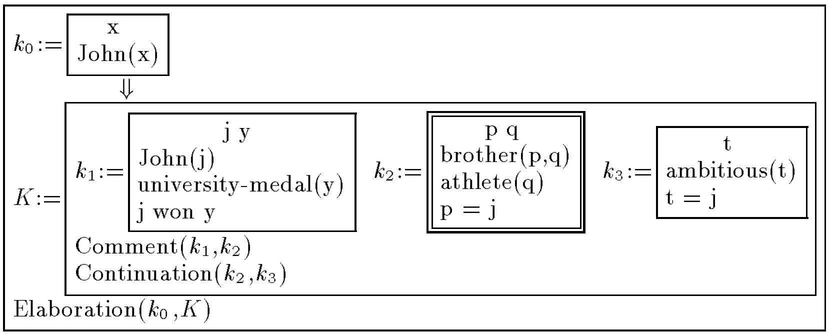

Figure 2. SDRS for (10)

In (10), *he* should preferably be resolved to John, not to his brother, as the material within dashes is parenthetical.[10] To deal with such sentences we have to modify the SDRS construction and take advantage of discourse structure. There should be a way to denote that the discourse referents introduced in the dashed sentence are parenthetical and are not preferred for further selection. We choose to use a double-framed box that denotes a lesser degree of preference for this purpose. The *Comment* relation implies that the dash-interpolated constituent is a digressive commentary and all of them form an *Elaboration* of the topic *John* (the topic is indicated by a downarrow). The relevant SDRS is given in Figure 2.

(11) a. He reported the decision: we were forbidden to speak with the chairman directly.

b. He reported the decision; we were forbidden to speak with the chairman directly. (Nunberg, 1990, p. 13)

(11a) (=(6a)) takes the decision to be the ban of spoken interaction with the chairman. (11b) (=(6b)), on the other hand, indicates that because of the ban the spokesman, not another person of the group, reported to the chairman. (11b) can also have another interpretation that the spokesman, not the chairman announced the decision as the others were forbidden to talk with the chairman. The first distinction can be captured by changing the SDRS building algorithms and directing the punctuation mark to the appropriate relation (the constituent l *elaborating* k in (11a) and *explaining* it in (11b)) as shown in Figure 3.[11] However, resolving such an ambiguity as in the multiple interpretations of (11b) without contextual information is a problem. More examples can be found in (Say and Akman, 1996b). We aim to extend our coverage to a fuller set of uses of various punctuation marks. After such

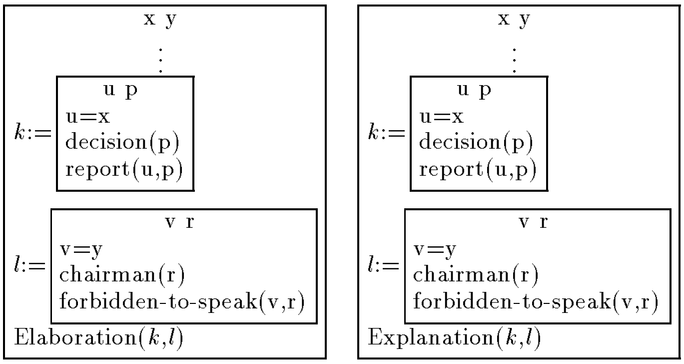

Figure 3. SDRSs for (11a) and (11b)

a treatment, we hope to make it a worthwhile endeavour to have the results apply in a computational setting. A preliminary study (Say and Akman, 1997) on dashes show (apart from preference anaphoric constructs) how dash usage indicates certain relations and how some dashed constructs imply intonational prominence within informational focus. The framework we are using may prove inadequate at certain points. Trying to indicate non-truth conditional constructs on top of a truth-conditional theory such as DRT may be problematic and limiting. Within the domain of higher-level marks (outside structural marks), a framework such as in (Pascual and Virbel, 1996) might prove a better approximation for a more complete theory.

7. Conclusion

Based on our discussion above, we can list the desiderata for a theory of punctuation. It should be a unified account of the syntactic, semantic, and discourse related effects of punctuation. It should account for both structural and text-level punctuation and be formal enough to be applied in the analysis and generation of written language. There is yet no such theory (except, maybe, for (Jones, 1997) but his is from a syntactic perspective) but the information-based perspective outlined in Section 6 seems promising and justified for covering the above multi-dimensioned criteria.

On the other hand, assuming that such a theory does eventually come into existence, we need more corpus-based studies observing current usage practices in different languages (including those with non-Latin scripts). For tuning the punctuation modules in NLP, more metrics of the sort punctuation-complexity (Jones, 1994b) and stability (Bayraktar, 1996) would be useful in characterizing specific kinds of texts.

Acknowledgements

We are indebted to three reviewers of *Computers and the Humanities* for thoughtful criticism and suggestions. Especially, the second referee had a major impact on the content and organization of the paper. We also thank Nancy Ide, Editor-in-Chief, for her encouragement. As usual, all the remaining inadequacies should be blamed on us.

The first author is grateful to the Scientific and Technical Research Council of Turkey (program code: TUBITAK-BAYG/NATO-A2) for financial aid.

Notes

[1] All cited works are mostly on English punctuation though there is some cross-linguistic work (Akram and Saadeddin, 1987; Simard, 1996; Twine, 1984).

[2] Refer, on the other hand, to (Henrichsen, 1995) for a highly radical interpretation of the notion of sentence. He develops a *Semicolon Grammar* based on a Categorial Grammar framework. His assumption is that language understanding is not restricted by sentence boundaries. The maximal projection in the grammar must be the discourse itself.

[3] One of the reviewers noted that this sentence is confusing in that there is an anaphoric reference from "both" back to "elaborative" or "contrastive", terms which are introduced parenthetically. We leave the sentence as is, as an example of an unintentional anaphoric implication of parentheticals.

[4] A planned extension of SUSANNE to spoken English is aimed at developing detailed definitions of speech notations (covering, among other things, discourse items and pauses).

[5] An exact comparison is not possible as they use different core grammars and Jones deletes 300 sentences of his data set because they are outside the coverage of his core grammar.

[6] An example is cue words, also known as *discourse markers* (Schiffrin, 1987), aiming to bring to the listener's attention the bond between the next utterance and the current discourse context.

[7] Coherence relations act as glue to the parts of a text by indicating implicit relations between those parts such that the content of one part may, for example, elaborate, exemplify, or explain that of another.

[8] See (Hoffman, 1995) which develops a grammar formalism that handles information structure of a "free" word-order language, Turkish, in parallel with predicate-argument structure.

[9] We accept that this sentence is contrived especially from an NLG perspective (as noted by one of the reviewers, it would be preferable to generate "Jane writes books on England, as do Joe and Sue.") but it is a clear example of the anaphoric aspects of punctuation.

[10] Nunberg (1990, p. 105) claims that parentheses (*not* dash-interpolation) completely constrain any referents inside serving as antecedents to external anaphors but we think that dashes can also serve such a role in a defeasible (not fully enforced) way.

[11] Dale (1991a; 1991b) suggests *Elaboration* and *Causation* respectively, as also pointed out by one of the reviewers.

References

ACL/DCI. Association for Computational Linguistics Data Collection Initiative, CD-ROM 1, 1991. http://www.ldc.upenn.edu

Akram, Mohammed and A. M. Saadeddin. "Target-World Experiential Matching: The Case of Arabic/English Translating". *Quinquereme*, 10(2):137–164, 1987.

Asher, Nicholas. *Reference to Abstract Objects in Discourse*. Kluwer, Dordrecht, Netherlands, 1993.

Bayraktar, Murat. *Computer-Aided Analysis of English Punctuation on a Parsed Corpus: The Special Case of Comma*. Master's thesis, Dept. of Computer Engineering and Information Science, Bilkent University, Ankara, Turkey, 1996.

Bolinger, Dwight. *Intonation and Its Uses: Melody in Grammar and Discourse*. Stanford University Press, Stanford, California, 1989.

Briscoe, Ted and John Carroll. "Developing and Evaluating a Probabilistic LR Parser of Part-of-Speech and Punctuation Labels". In *Proceedings of International Workshop on Parsing Technologies*, pages 48–58, Prague, Czech Republic, 1995.

Briscoe, Ted. *Parsing (with) Punctuation*. Technical report, Rank Xerox Research Centre, Grenoble, France, 1994.

Briscoe, Ted. "The Syntax and Semantics of Punctuation and Its Use in Interpretation". Pages 1–8. In (Jones, 1996a).

Chafe, Wallace. "Punctuation and the Prosody of Written Language". *Written Communication*, 5(4):395–426, 1988.

Cruttenden, Allen. *Intonation*. Cambridge University Press, Cambridge, UK, 1986.

Dale, Robert. "Exploring the Role of Punctuation in the Signalling of Discourse Structure". In *Proceedings of a Workshop on Text Representation and Domain Modelling: Ideas from Linguistics and AI*, pages 110–120. Technical University of Berlin, Berlin, Germany, 1991a.

Dale, Robert. "The Role of Punctuation in Discourse Structure". In *Working Notes for the AAAI Fall Symposium on Discourse Structure in Natural Language Understanding and Generation*, pages 13–14. Asilomar, CA, 1991b.

Doran, Christine. "Punctuation in Quoted Speech". Pages 9–18. In (Jones, 1996a).

Douglas, Shona and Matthew Hurst. "Layout and Language: Lists and Tables in Technical Documents". Pages 19–24. In (Jones, 1996a).

Ehrlich, Eugene. *Theory and Problems of Punctuation, Capitalization, and Spelling*. McGraw-Hill, Hong Kong, 1992.

Engdahl, Elisabet and Enric Vallduví. "The Linguistic Realization of Information Packaging". *Linguistics*, 34:459–519, 1996.

Fornell, Jan. "Punctuation in the Bravice English-to-Japanese Machine Translation System". Pages 25–32. In (Jones, 1996a).

Francis, W. Nelson and Henry Kučera. *Frequency Analysis of English Usage: Lexicon and Grammar*. Houghton Mifflin, Boston, MA, 1982.

Garside, Roger, Geoffrey Leech, and Geoffrey Sampson, editors. *The Computational Analysis of English*. Longman, London, 1987.

Grosz, Barbara J. and Candace L. Sidner. "Attention, Intentions, and the Structure of Discourse". *Computational Linguistics*, 12(3):175–204, 1986.

Hall, Nigel and Anne Robinson. The Punctuation Project. School of Education, Manchester Metropolitan University, Manchester, UK, 1996. http://bll.edu.aca.mmu.ac.uk/punctuation.html

Harris, Roy. *Signs of Writing.* Routledge, London, UK, 1995.

Henrichsen, Peter Juel. *Does the Sentence Exist? Do We Need It?* Unpublished Paper, Institute of Linguistics, University of Copenhagen, Copenhagen, Denmark, 1995.

Hoffman, Beryl. "Integrating 'Free' Word Order Syntax and Information Structure". In *Proceedings of the 1995 Conference of the European Chapter of Association for Computational Linguistics*, pages 245–252, Dublin, Ireland, 1995.

Humphreys, Lee. "Book Review: The Linguistics of Punctuation". *Machine Translation*, 7:199–201, 1993.

Ince, Bahar. *Punctuation: The Special Case of Comma Categories.* Senior Project Report, Dept. of Computer Engineering and Information Science, Bilkent University, Ankara, Turkey, 1996.

Jackendoff, Ray. *X-bar Syntax: A Study of Phrase Structure.* MIT Press, Cambridge, MA, 1977.

Jones, Bernard. Can Punctuation Help Parsing? Acquilex-II Working Paper 29, Computer Lab., Cambridge University , Cambridge, UK, 1994a.

Jones, Bernard. "Exploring the Role of Punctuation in Parsing Natural Language". In *Proceedings of the 15th International Conference on Computational Linguistics (COLING '94)*, pages 421–425, Kyoto, Japan, 1994b.

Jones, Bernard. "Exploring the Variety and Use of Punctuation". In *Proceedings of the 17th Annual Cognitive Science Conference*, pages 619–624, Pittsburgh, PA, 1995.

Jones, Bernard, editor. *Punctuation in Computational Linguistics*, UCSC, Santa Cruz, CA. SIGPARSE 1996 (Post Conference Workshop of ACL96). Available from Human Communication Research Center, University of Edinburgh, UK, 1996a. http://www.cogsci.ed.ac.uk/hcrc/publications/wp-2.html

Jones, Bernard. "Towards a Syntactic Account of Punctuation". In *Proceedings of the 16th International Conference on Computational Linguistics (COLING '96)*, pages 604–609, Copenhagen, Denmark, 1996b.

Jones, Bernard. "Towards Testing the Syntax of Punctuation". In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics-Student Session*, pages 363–365, Santa Cruz, CA, 1996c.

Jones, Bernard. *What's the Point? A (Computational) Theory of Punctuation.* PhD thesis, Centre for Cognitive Science, University of Edinburgh, Edinburgh, UK, 1997.

Kamp, Hans and Uwe Reyle. *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory.* Kluwer, Dordrecht, Netherlands, 1993.

Karlsson, Fred, Atro Voutilainen, Juha Heikkila, and Arto Antilla, editors. *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text.* Mouton de Gruyter, Berlin, Germany, 1994.

Lee, Sherman. *A Syntax and Semantics for Text Grammar.* Master's thesis, Engineering Dept., Cambridge University, Cambridge, UK, 1995.

Levinson, Joan Persily. *Punctuation and the Orthographic Sentence: A Linguistic Analysis.* PhD thesis, Dept. of Linguistics, City University of New York, NY, 1985.

Mann, William C. and Sandra A. Thompson. *Rhetorical Structure Theory: A Theory of Text Organization.* Technical Report RS-87-190, USC Information Sciences Institute, University of Southern California, Marina Del Rey, CA, 1987.

McDermott, John. *Punctuation for Now.* MacMillan, Hong Kong, 1990.

Meyer, Charles F. *A Linguistic Study of American Punctuation.* PhD thesis, University of Wisconsin-Milwaukee, WI, 1983.

Meyer, Charles F. "Punctuation Practice in the Brown Corpus". *ICAME Newsletter*, pages 80–95, 1986.
Meyer, Charles F. *A Linguistic Study of American Punctuation.* Peter Lang, New York, NY, 1987.
Min, Young-Gie. "Role of Punctuation in Disambiguation of Coordinate Compounds". Pages 33–40. In (Jones, 1996a).
Muskens, Reinhard. "Combining Montague Semantics and Discourse Representation". *Linguistics and Philosophy*, 19:143–186, 1996.
Nunberg, Geoffrey. *The Linguistics of Punctuation.* Number 18 in CSLI Lecture Notes. CSLI Publications, Stanford, CA, 1990.
Oehrle, Richard T. Lecture Notes: Prosody, Information and Grammatical Architecture. *Seventhth European Summer School in Logic, Language and Information*, Barcelona, Spain, 1995.
Osborne, Miles. "Can Punctuation Help Learning?" In Stefan Wermter, Ellen Riloff, and Gabriele Scheler, editors, *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*, Lecture Notes in Artificial Intelligence, Number 1040, pages 399–412. Springer-Verlag, Berlin, 1996.
Parkes, M. B. *Pause and Effect: An Introduction to the History of Punctuation in the West.* University of California Press, Berkeley, CA, 1993.
Partridge, Eric. *You Have a Point There: A Guide to Punctuation and its Allies.* Routledge, London, UK, 1993.
Pascual, Elsa and Jacques Virbel. "Semantic and Layout Properties of Text Punctuation". Pages 41–47. In (Jones, 1996a).
Pereira, Fernando and David Warren. "Definite Clause Grammars for Language Analysis — A Survey of the Formalism and a Comparison with Augmented Transition Networks". *Artificial Intelligence*, 13(3):231–278, 1980.
Prevost, Scott and Mark Steedman. "Specifying Intonation from Context for Speech Synthesis". *Speech Communications*, 15:139–153, 1994.
Sampson, Geoffrey. "Book Review: The Linguistics of Punctuation". *Linguistics*, 30(2):467–475, 1992.
Sampson, Geoffrey. *English for the Computer: The SUSANNE Corpus and Analytic Scheme.* Oxford University Press, Oxford, UK, 1995.
Say, Bilge and Varol Akman. "Information-Based Aspects of Punctuation". Pages 49–56. In (Jones, 1996a).
Say, Bilge and Varol Akman. "An Information-Based Treatment of Punctuation in Discourse Representation Theory". In *Second International Conference on Mathematical Linguistics*, Tarragona, Spain, 1996b.
Say, Bilge and Varol Akman. Dashes as Cues to Discourse Structure. Manuscript. Dept. of Computer Engineering and Information Science, Bilkent University, Ankara, Turkey, 1997.
Say, Bilge. *An Information-Based Approach to Punctuation.* PhD Proposal, Dept. of Computer Engineering and Information Science, Bilkent University, Ankara, Turkey, 1995. http://www.cs.bilkent.edu.tr/˜say/bilge.html
Schiffrin, Deborah. *Discourse Markers.* Cambridge University Press, Cambridge, UK, 1987.
Scholes, Robert J. and Brenda J. Willis. "Prosodic and Syntactic Functions of Punctuation — A Contribution to the Study of Orality and Literacy". *Interchange*, 21(3):13–20, 1990.
Shiuan, Peh Li and Christopher Ting Hian Ann. "A Divide-and-Conquer Strategy for Parsing". Pages 57–66. In (Jones, 1996a).
Simard, Marthe. "Considerations on Parsing a Poorly Punctuated Text in French". Pages 67–72. In (Jones, 1996a).
Smith, Carolena L. "Attitudinal Study of Graphic Computer-Based Instruction for Punctuation". *Journal of Technical Writing and Communication*, 3:267–272, 1986.

Srinivasan, V. "Punctuation and Parsing of Real-World Texts". In K. Sikkel and A. Nijholt, editors, *Proceedings of the Sixth Twente Workshop on Language Technologies*, pages 163–167, Enschede, Netherlands, 1991.

Steedman, Mark. "Structure and Intonation". *Language*, 67(2):260–296, 1991.

Taylor, Lita J. and Gerry Knowles. *Manual of Information to Accompany the SEC Corpus*. University of Lancaster, Lancaster, UK, 1988.

Twine, Nanette. "The Adoption of Punctuation in Japanese Script". *Visible Language*, 18(3):229–237, 1984.

Vallduví, Enric. *The Informational Component*. Garland, New York, 1992.

White, Micheal. "Presenting Punctuation". In *Proceedings of the Fifth European Workshop on Natural Language Generation*, pages 107–125, Leiden, Netherlands, 1995.

Authors' Vitae

B. Say

Bilge Say received her BS in Computer Engineering from Middle East Technical University, Ankara, Turkey, in 1990, and her MS in Computation from Oxford University, Oxford, UK, in 1991. She worked two years as a systems support engineer in the industry. Currently, she is a PhD student at Bilkent University, Ankara, Turkey, studying the information-based aspects of punctuation. She has recently conducted research at the Computer Lab of Cambridge University, Cambridge, UK, on an extended visit.

V. Akman

Varol Akman is a professor of computer engineering at Bilkent University, Ankara, Turkey. From 1980 to 1985, he was a Fulbright scholar at Rensselaer Polytechnic Institute, Troy, New York, where he received a PhD degree in computer engineering. Prior to joining Bilkent in 1988, he held a senior researcher position with the Centrum voor Wiskunde en Informatica, Amsterdam, the Netherlands. His current research areas include artificial intelligence models of context, computational aspects of situation theory, and in general, language and philosophy.

Address for correspondence: Varol Akman
Dept. of Computer Engineering and Information Science,
Bilkent University,
06533 Bilkent, Ankara, Turkey
akman@cs.bilkent.edu.tr