The Rise of Text Linguistics

Dilafruz Izatullayevna Xodjaeva¹ and Juraeva Guzal Jabbor Kizi²

¹Head of ESP for natural subjects department
²Master of English linguistics specialty
Bukhara State University

Abstract – A decisive step in the rise of modern ‘Western Continental’ linguistics was to instate the central tenet that language should be considered a uniform, stable and abstract system apart from variations due to individual or social contexts and activities.

Keywords – Linguistics, Social Contexts, Scope of Phonemics and Morphemics, Sentence Types.

I. INTRODUCTION

The description of the language should be couched in highly general statements about the language as a whole or even in universal ones about all languages. This (Saussurean) programme inaugurated the search for the complete set of constraints upon ‘language in and for itself’ (‘langue’) without reference to the use of language (‘parole’). The constraints should cover all valid instances or sets of instances of the language.

Instead of accepting the central tenet as an incontestable a priori principle, as ‘mainstream’ linguistics has widely done, we might view it as a working hypothesis that should be tested. The crucial tests would surely be the convergence among the data discovered and described, and the consensus among linguists about the constraints that govern the data. Beyond the scope of phonemics and morphemics, however, these tests have not been met. Instead, we can observe linguists choosing between two compensatory strategies for tapping constraints that had not yet been properly discovered and stated. Either they have sought to describe corpuses of recorded real data that are ‘naturally’ constrained by their contexts of occurrence, or they have used their own intuition and introspection as a source of constraints while inventing language data, mostly isolated sentences. The second strategy looks easier than the first but turns out not to be because it obliges linguists to devise artificial constraints to describe some ‘well-formed’ sentences and to exclude some ‘ill-formed’ ones that would be unlikely to appear in real data. Not surprisingly, the switch from real corpuses over to invented sentences led to an increasing breakdown in convergence and consensus.

The situation was obscured somewhat by the versatile role of the sentence. For Saussure (1966 [1916]: 124), the sentence was “the ideal type of syntagma, but it belongs to speaking [parole], not to language [langue]”. For Bloomfield (1933: 170ff), the sentence was “an independent form not included in any larger (complex) linguistic form”; “perhaps all languages distinguish” “sentence types”. For Chomsky (1957: 55 and 1965: 34), the sentence was a main “notion that must be defined by general linguistic theory”, and the unit to which “a descriptively adequate grammar … assigns structural descriptions … in accordance with the linguistic competence of the native speaker”.

Corresponding Author: Dilafruz Izatullayevna Xodjaeva
II. LITERATURE REVIEW

For Halliday & Hasan (1976: 8, 244, 235), a sentence was “the highest unit of grammatical structure” and was defined as “any set of clauses that are hypotactically and/or paratactically related . . . sentences cannot be rearranged, as a coordinate structure can, in different sequences”. This versatility suggests that the sentence provided a strategic compromise between language versus use by being both (1) a formal or ‘grammatical’ pattern whose obligatory or optional constituents are specified by the general language system, and (2) a fairly self-contained utterance pattern. Every reasonable sentence, real or invented, entails some constraints reflecting its possible uses. So the sentence has readily been accepted as the standard realisation of semantic and pragmatic entities like the ‘proposition’, ‘predication’, ‘utterance’, ‘assertion’, or ‘speech act’.

Yet even so versatile a notion of the sentence has not supplied linguistic theories and models with any definitive set of constraints, either for corpuses of real data or for accumulations of invented data. Naturally, some linguists looked ‘beyond the sentence boundary’, giving a vital impulse to text linguistics in the late 1960s and early 1970s. For the sake of continuity, early text linguists based the ‘text’ in more abstract or concrete ways on the sentence, witness the tactic of studying the constraints only within ‘sentence pairs’, especially question and answer, the grammatical structure of the preceding sentence being treated as a minimal context (e.g. Isačenko 1965).

III. ANALYSIS

To the degree that the ‘sentence’ was officially a ‘grammatical’ or ‘syntactic’ unit, research was dominated by notions like ‘text grammar’ (e.g. van Dijk 1972), ‘text syntax’ (e.g. Dressler 1970), ‘hyper-syntax’ (Palek 1969), and ‘macro-syntax’ (Gülich 1970), which allowed some double-tracking by referring to the sentence while moving ‘beyond’ it. Research sought to project the prevailing concepts and methods from sentences over to texts with relatively modest revisions, yet justified text studies by raising grammatical issues not well accounted for within the single sentence. For instance, the ellipsis found in “faithful recordings of spontaneous conversation” could be grasped as “agrammatical connection among sentences” rather than a phenomenon inside the single sentence, yet can be constrained by deriving each instance from an “underlying sentence [that] native speakers reconstruct” (Gunter 1963: 137, 140).

Such double-tracking at the boundaries of the sentence eventually led to a reassessment of its role in ‘grammar’. In a genuine ‘text grammar’, the sentence could be neither the ‘largest linguistic form’ to be studied (Bloomfield) nor the only unit to which ‘structural descriptions’ are assigned (Chomsky). Such a ‘grammar’ was expected also to “specify the formal features that make a sequence of sentences into a text” (Isačenko 1965: 164); or to “distinguish whether two successive sentences belong to the same text” (Dressler 1970: 66), or to “specify the formal properties determining a typology of texts” (van Dijk 1972: 5); or to “explain how speakers of a language […] are able to distinguish between grammatical and ungrammatical texts” (van Dijk 1972: 4); and so on.

But no such ‘text grammar’ managed to supply adequate constraints. Non-grammatical factors were required even to account for manifestations that are undeniably ‘grammatical’, such as pronouns (e.g. Steinitz 1971), articles (e.g. Weinrich 1969) and tenses (e.g. Weinrich 1964); and the same holds far more plainly textual factors, such as cohesion, coherence, topic progression, narrativity, and situationality, which also affect grammar. So text linguistics energetically explored semantic and pragmatic issues, including those previously discounted for describing abstract, isolated sentences.

This exploration engendered a wide spectrum of definitions for the text. For Isenberg (1970: 1), a text was a “coherent sequence of sentences”. For Harweg (1968a: 148), a text was “a series [Nacheinander] of linguistic units constituted by uninterrupted pronominal chaining”. For van Dijk, a text was “the abstract theoretical construct underlying discourse” (1976: 3); or “a sentence sequence with a macro-structure”, i.e. a “global structure of meaning” (1978: 41; cf. van Dijk 1979). For Hartmann (1968: 220), the text was a “finite ordered set of textually complicated partial signs of various sorts and functions” and hence “the linguistically significant originary sign of language”. For Weinrich (1976: 186), a text was “an ordered sequence of language signs between two noticeable discontinuations [Unterbrechungen] of communication”. For Halliday & Hasan (1976: 1, 4, 23) a text was “any passage, spoken or written, of whatever length, that does form a unified whole”; it was variously said to be ‘defined’ by ‘relations of meaning’ or by ‘cohesion’ and ‘register’. For Schmidt (1973: 144ff), a text was “a linguistic realisation of textuality as an organisational form for the formation of pluralistic complexes”, hence a “matrix for the mutually relevant and coherent occurrence of elements in the language system”.

This brief gallery of text definitions in text studies was emblematic for the lively discussions of fundamental concepts and problems merely implicit in ‘sentence
linguistics’, as earlier approaches now came to be called. If the text is so diverse an entity and yet consists of sentences, then the sentence should also have a high diversity that has hitherto not been acknowledged or strictly formalised, but has remained inherent in the intuitions of linguists or in the conditions of social contexts. Some of this diversity was ambivalently acknowledged at times, e.g. when ‘sentence linguists’ sought richer constraints by adding a ‘semantic component’ but balked at advancing it from an ‘interpretative’ to a properly ‘generative’ role, doubtless fearing to lose the formality; the ‘syntactic features’ and ‘logical form’ proved even less conducive to convergence and consensus than did purely ‘syntactic’ notions.

When the text became an object of ‘syntactic’ inquiry, the pressure redoubled to reintroduce the semantic and pragmatic constraints that had previously been diluted by discounting real corpuses of empirical data and centring linguistic theory on the syntax of the single sentence. Thus, Weinrich (1967: 112) redefined ‘syntax’ as a ‘dialectic’ that ‘investigates all semantic elements of language in their relation to the speech situation’. Dressler’s (1970: 68) ‘text-syntax’ ‘favoured hypotheses that take account of the actualities of the speech process’. Gülich’s (1970: 46, 297) ‘macro -syntax’ eschewed the concept of ‘sentence’ as “highly questionable however you define it”, and looked for the ‘arrangement signals’ (‘GliederungsSIGNALE’) marking “the beginning or end of spoken texts or text segments”.

Daneš’s (1970) model of ‘thematic progression’ in texts reflected his well-known ‘three-level approach to syntax’, wherein the ‘grammatical structure of sentence’ was paralleled by its ‘semantic structure’ and its pragmatic ‘organisation of utterance’ (1964: 226ff). And Petöfi’s (1978: 40) ambitious “text-structure/world structure theory (…) aims at describing the syntactic, semantic, and pragmatic structures of natural language text” through a “formal semiotics”.

It was high time to consider whether the text would belong on the side of language or of language use. The simplest ‘mainstream’ recourse of putting it among language use had the drawback that linguistic theory had not just neglected use, but had described it in negative terms to rationalise its neglect. For Saussure (1966 [1916]: 13, 15, 19, 11), ‘parole’ was “individual”, “heterogeneous”, and “momentary”; “speech cannot be studied … for we cannot discover its unity”. For Chomsky (1965: 4, 44, 58, 201), ‘actual speech’ in ‘performance’ was “scattered” and “deficient” and had a “degenerate quality”, replete with “false starts … fragments, and deviant expressions”.

Text linguists in turn remained cautious about putting the text in the domain of actual use even though they viewed use much more positively. Van Dijk (1972: 3ff) reasoned: “although theories, and especially grammars, do not ‘directly’ describe empirical phenomena, but only formally reconstruct the entities, relations and systems ‘underlying’ these phenomena, it is necessary to postulate entities or regularities which are as close as possible to the empirical phenomena without losing their general character”. So he proposed to make the text a “theoretical concept” and an “abstract unit [that] underlies a meaningful utterance [of a] discourse”; “any utterance is the token of a text” (1978: 41, 1972: 317ff, 1976: 3). Similar reasoning motivated such conceptions as Bellert’s (1970: 366) “idealised text with no digressions … expressing one uniform reasoning, continuous plot, etc.”; Dressler’s (1970: 65) “potential text in the sense of a pre-linguis- tic text”; Schmidt’s (1973: 150) “meta-linguistic “text formulary”, being “a set of linguistic constituents ordered by a thematic deep structure”; and - Harweg’s (1968a: 152, 1971: 179) “emic text defined by text- immanent criteria” as opposed to “etic text determined by extra-linguistic criteria not based on language structure”. Like Saussure’s ‘speech’, Harweg’s ‘etic’ side lacked “unifying delimiting criteria”, witness the “etic text beginnings [of] novels, short stories, poems, scientific treatises”; still, he allowed that emic and etic texts might coincide to varying degrees (1968a: 152ff, 1968b: 348).

For Hartmann, in contrast, “sentence formation proceeds through rules of individual languages”, whereas “text formation proceeds through principles applying to all languages and through extralinguistic stylistic and individual requirements”, e.g. “motives, intentions, effects” (1968: 218ff). This formulation justified treating the sentence within a ‘grammar’ in a ‘language’ perspective yet suggested that speakers who ‘form’ a text are transcending the boundary between langue and parole plus the boundary between the particular language and ‘all languages’. The text would be both more general and more specific than the sentence, and text theory would need to span the whole range from abstract ‘universals’ over to concrete situational factors.

Hartmann’s vision signalled his remarkable agility as a theorist, which powerfully influenced his students (e.g. Harweg, Koch, Schmidt, Rieser, Wienold, de Beaugrande). Like the British functional school that did not call itself ‘text linguistics’ (e.g. Firth, Halliday, Sinclair, and their students), he realised that separating the text from the communicative context and situation obscures rather than clarifies the question of what a theory should account for. So we should not confine the term ‘text’ to a theoretical abstraction as...
IV. DISCUSSION

Today, the ‘text’ is widely defined as an empirical communicative event given through human communication rather than specified by a formal theory. Each such event ‘rides on’ a dynamic dialectic between the ‘virtual system’ of language (the repertory of possibilities) and the ‘actual system’ constituted by the choices of the text producer; the text is thus on neither side of language versus use, but integrates and reconciles the two. Our task is to describe, as empirically and realistically as we can, the processes whereby communicative participants can and do produce and receive texts. The task plainly demands interdisciplinary research between text linguistics and psychology, sociology, ethnography, and so on, all of whom work with real data from the standpoint of human activities.

Defining the ‘text’ and our main task this way brings ‘text linguistics’ into line with the research tradition of discourse analysis, which has operated ‘bottom up’ with realistic data obtained from fieldwork. Since its earliest formulations (e.g. Harris 1952), investigators have looked a real data beyond the sentence and developed a ‘grammar’ incorporating semantic and pragmatic issues, such as how different languages deploy grammatical resources for expressing events and participants in narratives (see now Longacre 1990). The ethnographic emphasis obviated the need for the abstract theoretical formalisations cultivated in ‘mainstream linguistics’, including the isolation of ‘language in and for itself’ from the use of language (cf. Stubbs 1993; Sinclair 1994). Today, the same emphasis has brought discourse analysis into the mainstream, while formalist linguistics drifted into a ‘scientific crisis’ by failing to meet its announced goals, notably to formulate a complete ‘generative grammar’ for any natural language. The constraints once excluded on theoretical grounds now had to be reintroduced, and discourse analysis steadily merged with text linguistics.

The latest and most decisive step has been the compilation of the ‘Bank of English’, a vast computerised corpus of contemporary spoken and written English discourse, which enables reliable statements about the typical patterns and collections of English (Sinclair 1992a, 1992b; Baker et al. eds. 1993). Some of the many regularities that are not open to mere intuition or introspection emerged at 20 million words (1992 figure); others at 200 million words (1994 figure). For example, the verb brook was found to have a clearly defined set of constraints on its usage (Sinclair 1994). Grammatically, it takes negation (usually not just before it or, no just after), and does not appear with the first or second person singular as the subject. Semantically, its direct object is a concept associated with opposition, interference, or delay. Pragmatically, the subject must be in a position of sufficient authority to carry the performative perlocutionary force entailed in declaring what it will not brook.

The implications Sinclair (1994) draws from such data are likely to be felt through all of linguistics. (He provocatively compares the large corpus to the telescope in astronomy and the microscope in biology – instruments which vastly enhanced the scientific status of inquiries). First, the rift between ‘langue’ versus ‘parole’ is finally healed, because we can derive the regularities in the language fairly directly from real data. Second, the pendulum swings decisively back toward convergence and consensus because such constraints as those upon the verb brook are strongly implicated by the totality of data in the display, whereas the constraints upon a single invented sentence that might hardly ever be uttered can be endlessly disputed. Third, the semantic and pragmatic constraints set aside by formalist models are effectively restored and placed on a more realistic basis than ever before. Fourth, we see that the appropriate unit for analysis is not the morpheme, word, or phrase but the ‘co-text’, the utterance environment, of variable lengths, in which words of phrases tend to appear together. And fifth, we no longer need to cut up linguistics into tidy ‘levels’ or ‘components’ for separate analysis and description; even the staid division between grammar and lexicon proves to be merely two complementary points of view upon the same phenomena.

Text linguistics too will feel the impact. First, the dialectic between virtual system and actual system becomes far clearer and more direct when the function of a word of phrase in the representative corpus can be explicitly correlated with the function it has in a given text. Second, the text as whole does not need a formal circumscribed definition; for most purposes, it can be perfectly well grasped as one actualization within a set of attested cross-sections we can pull up for display as we choose. Third, Sinclair (1994) suggests that the displays show competence if read vertically and performance if read horizontally; but this ‘competence’ would at most be probabilistic because speakers of English know how to produce data that have not yet occurred in the corpus, although they may well not be likely to the sentence...
boundary loses its theoretical significance and becomes an empirical phenomenon which, like the other boundary signals of punctuation or intonation, occurs with higher probability after some word combination (Firth’s ‘collocations’) and grammatical patterns (Firth’s ‘colligations’) than others. Fourth, the production of a text is viewed as a set of interactive local choices, some of which are made well before the key word appears in the sequence. And fifth, we can use subcorpora as a data-driven bottom-up method for characterising text types and discourse domains, a task for which speculative schemes have remained ineffectual.

We can now reach a consensus about realistic definitions of our two main terms: text being the individual communicative event within the ongoing interaction of discourse. Our shared research interests should decide how abstract or concrete, how theoretical or empirical, these terms should be in specific projects with explicit goals, e.g. studies of ‘discourse for special purposes’ (DSP) in technical fields (an analogue to ‘register’ studies) with the goal of developing strategies to make specialised knowledge more generally accessible. The same interests could resolve other thematic questions in text linguistics that are not resolvable on general theoretical grounds, e.g. how far we would analyse a text in an ‘ascendant mode from smaller units to the text or in a descendant mode from text to its constituent units’ (Kallmeyer et al. eds. 1971: 57).

Text research is also swinging into line with discourse processing, an interdisciplinary exploration of the processes that make discourse both feasible useful in so many settings (e.g. Freedle ed. 1977, 1979; Beaugrande 1980, 1984; van Dijk & Kintsch 1983) and a major sector in cognitive science, itself a convergence among psychology, computer science, sociology, and anthropology, along with their offshoots and intermediaries like psycholinguistics, cognitive psychology, artificial intelligence, sociolinguistics, and social psychology. The question of ‘ascendant’ versus ‘descendant’, for instance, could be assessed in light of finding that text receivers combine both tactics, doing ‘bottom-up’ scans of the local text units while forming and testing ‘top-down’ hypotheses about the global organisation, e.g. the ‘macro-structures’ of overall coherence.

Further, text linguistics has become strongly engaged with pragmatics (e.g. Schmidt 1973, 1976). Instead of giving a set of ‘rules’ that “turn these objects into acts” and “place these acts in a situation” (van Dijk 1976: 190), pragmatics investigates how texts and co-texts (not sentences) are always already acts in situations, even while the syntactic or semantic operations are being carried out. Hence, pragmatics forms the outermost framework for approaching syntax and semantics in terms of communicative activities.

And finally, text linguistics has assumed a cordial relation to stylistics (e.g. Enkvist 1977), which mainstream linguistics tended to marginalise by situating it in the use of language rather than in language itself. For Weinrich (1969: 69), ‘the border’ consists of text linguistics exploring the syntactic function of an element such as the indefinite article in the language, i.e. “in every real or possible text”, whereas ‘stylistics’ “explores the particular use in a certain text” by one ‘author’; in between lies ‘text typology’, stating for which “text types certain affinities can be determined”. Here, we may feel reminded of Weinrich’s (1967: 112) already cited definition of ‘syntax’ as a ‘dialectic’ that “investigates all semantic elements of language in their relation to the [pragmatic] speech situation”; the relation between the function of an element in the language and in a given text is surely dialectical too. As with Hartmann and Sinclair, the old static opposition between ‘language’ and ‘parole’ was here given a dynamic reinterpretation making the text the entity that interfaces the two sides.

The increasing engagement of text linguistics with the four domains just cited discourse analysis, discourse processing, pragmatics, and stylistics has both increased our flexibility and guided us in resetting our priorities. We have grown more aware of how to coordinate theoretical with empirical, and categories with data, when we set up and implement a research project. Instead of following one set of orthodox all-purpose procedures, we seek theoretical principles for telling which to address. Moreover, we acknowledge that, as a communicative event, the text does not simply hold still like an inert object while we analyse it. Instead, we become engaged in ‘retexualising’ it, which can influence the conditions of its ‘textuality’. We must consider how far our intentions and interpretations as investigators parallel, or differ from, those of the discourse participants whose activities are now expressly included in the scope of our studies, e.g. how far we may develop more elaborate representations to register and resolve multiple meanings, or single out deviations or disturbances, well beyond what would reach the consciousness of a speaker or hearer. Such issues too can be empirically approached. In retrospect, then, we might distinguish three stages in the rise of ‘text linguistics’. The earliest stage (late 1960s, early 1970s) was dominated by ‘text grammar’ and remained fairly circumscribed and uniform by working to extend prevailing theories from sentence to text. The next stage (late 1970s, early 1980s) was more dominated by ‘textuality as a structure’ with ‘both linguistic and social aspects’. The
The Rise of Text Linguistics

1. The text is not merely a linguistic unit, but an event of human action, interaction, communication, and cognition. The notion of the text being merely an instance of language has a long tradition both in folk-wisdom and in the various specialised disciplines that have studied texts – from ancient grammar and rhetoric up through modern linguistics. The notion was hardly challenged as long as the ‘text’ did not have the status of a central or explicitly defined term. When language studies reached the point in the search for constraints where the ‘text’ did attain this status, its ‘multi-media’ qualities as a social and interactional event came into focus. The perennial question, ‘Is it still linguistics?’ generated during half a century of self-conscious questing for ‘language by itself’, seems quaint or irrelevant now. The constraints bearing upon texts must be comprehensively explored, whatever their sources.

2. The main source of data is naturally occurring texts and discourses. Through history, most studies of language have expropriated language data without expressly justifying its authenticity. Most samples in earlier studies were in practice taken from prestigious sources such as literary or sacred texts, whose specialised qualities were not considered problematic. Modern linguistics recognised the problems and emphatically turned to ordinary speech but, somewhat contradictorily, proposed to describe it apart from texts. The proposal eventually led to descriptions based on invented data, which was artificial in many respects including even its supposed ‘ordinariness’. Today, large corpuses of real data reveal the arguments in favour of this tactic to be either obsolete, e.g. that we cannot get our hands on enough data; or spurious from the start, e.g. that speakers have some idealised ‘competence’ whose relation to what they actually say is deeply uncertain. The corpuses also show that the supposed ‘uniqueness’ of utterance impedes us not to retreat from ‘surface structure’ to ‘deep structure’ but rather to exploit the commonalities of collocation and colligation that we can derive from the data.

3. Text analysis is rich and expansive. The amount of potentially relevant data for describing a text as a communicative event is in principle open. We can be content when we have uncovered some non-obvious and useful insights about the human consequences of the text, e.g. how and how well it makes knowledge accessible to people who need it.

4. In modern-day text research, the investigator has to engage and re-engage with the texts rather than strive for an idealised separation of subject from object or scientist from data.

5. The motives for doing text research in terms of epistemological interests bearing on the relations between texts and society have become more and more important for researchers in the field. Thus, e.g. projects addressing issues of inequality in the access to knowledge through discourse are socially relevant and thus enhanced. ‘Critical’ text
research should ultimately try to establish freedom of access to knowledge and to reveal and redistribute communicative power structures.

6. Text linguistics has an interdisciplinary perspective; linguistics cannot be the sole basis for a science of text and discourse. To understand a phenomenon as complex as language, we need to regard it from continually alternating and evolving viewpoints.

These issues have increased in importance within text linguistics. They all go in the direction of Coseriu’s (1971: 189) vision of “a new dimension” wherein “the text is a perspective concerning all of linguistics, and every linguistic discipline would be reconceived from this standpoint”.

REFERENCES

