EDITORIAL

The summer has more or less arrived and here is the Bulletin full of interesting reading to enjoy out in the sun. This issue of the Bulletin is a wonderful example of the span that the history of the language sciences entails. We have articles as different as one treating ancient Greek grammar and linguistics and one investigating the rather new phenomenon of (computational) corpus linguistics. We also have an article treating one of the major figures in the history of Swedish linguistics, Adolf Noreen, and his concept of the sememe.

It is nice to be able to give the readers such a wide range of subjects in one issue. However, every now and then we also believe it can be good to concentrate on one specific topic. Hence there are some special editions coming up in the near future. Furthermore, readers with ideas for special issues are invited to contact me, since we are always looking for new ideas.

Readers will see that articles have in the last few issues started to get longer, and I would like to clarify that we welcome **both** short and long articles. The exact word limit is decided in each case by the editorial board based on the reviewer(s)'s comments.

I would also like to take this opportunity to welcome two new members on the editorial board, Nicola McLelland and Fiona Marshall. Nicola and Fiona have worked with me on this issue as assistant editors, giving me their views on the articles that have come in and also helping out with proof reading. They have been of wonderful help and the next issue will be produced with Nicola in charge and Fiona and I assisting her since I will then be on leave, but contributors should still send materials to me, preferably at the following email address: a.m.t.tiedemann@rug.nl.

We hope this expansion of the editorial board will help to improve the standard of the Bulletin and hopefully it will also help us make sure it appears on time. With a team of four (including Richard, who is our Review Editor), between us we can discuss various ideas and hopefully come up with the best ones.

Unfortunately, this issue contains no reviews at all, which is something Richard and I are very sad about. We would like to encourage people to get in touch with us about the books in the publications received sections of this and previous Bulletins if there is something you would like to read and would consider reviewing. We know that many of you are interested in reading about the new books that have appeared in our field but, unfortunately, this is not possible unless you are also willing to review things.

Therese Lindström Tiedemann, *Groningen*
Editor

Note from the Vivian Law Prize Committee

It is a great pleasure to be printing the first recipient of the Vivien Law Prize in this issue of the Bulletin. The quality of the applications was extremely high, but Casper de Jonge's paper on Dionysius of Halicarnassus and the historiography of linguistics is a worthy winner. Given Vivien Law's own interest in both the history of linguistics and the writing of that history, it is fitting that the first prize-winner should bring the two together in this original way.

Dionysius of Halicarnassus as a Historian of Linguistics
The history of the theory of the 'parts of speech' in
De compositione verborum 2

Casper de Jonge
Leiden University, The Netherlands

1. Introduction

*P*artes orationis quot sunt? 'How many parts of speech are there?' It is with this question that the Roman grammarian Donatus (who was active around 350 AD) starts his *Ars Minor*. His answer is, of course: *octo*, 'eight'. Traditionally, we learn that the system of eight word classes, which we also find in the works of Apollonius Dyscolus (2nd century AD) and in the *Technê grammatikê* (3rd or 4th century AD), was the result of a long cumulative process: Plato (427-347 BC) identified two parts of speech, Aristotle (384-322 BC) three or four, the philosophical school of the Stoics (which was founded ca 315 BC) five or six, and Aristarchus (2nd century BC) and Dionysius Thrax (170-90 BC) eight (cf. Sluiter, 1998: 24-25).[1] This presentation of the history of the word class system has been criticised in recent years, but it is characteristic for the traditional historiography of linguistics, represented by scholars like Steinthal (1863), Benfey (1869) and Robins (1967 and later).[2] However, as far as we know, the first text that presented the history of the word class system in this way is Dionysius of Halicarnassus's work *On Composition* (*De compositione verborum*) (see Taylor, 1986: 177). Dionysius was a teacher of rhetoric, who was active in Rome between 30 and 8 BC. In this paper, I intend to make clear that the second chapter of Dionysius's *On Composition* can be considered the prototype of the traditional western approach to the history of linguistics.[3]

Dionysius of Halicarnassus published a number of rhetorical and critical treatises, in which he analysed the styles of classical Greek orators and historiographers. His work *On Composition* can be characterised as a synthesis of rhetorical, poetical, grammatical, metrical, musical, and philosophical views from earlier periods: Dionysius makes use of all these different disciplines in order to determine criteria for an effective composition. In the second chapter of this work, Dionysius discusses the history of the theory of τὰ μέρη τοῦ λόγου (*ta merê tou logou*), the 'parts of speech', which we would call 'word classes' (on these terms, see

---

[1] The Greek system of eight parts of speech (μέρη λόγου) consisted of
ὄνομα, ῥῆμα, μετοχή, ἄρθρον, ἀντωνυμία, πρόθεσις, ἐπίρρημα and σύνδεσμος. The Romans substituted the interjection for the article, thus listing the following eight word classes: *nomen, verbum, participium, pronomen, praepositio, adverbium, coniunctio, interiectio.*

[2] For objections to the traditional presentation of the history of the word class system, see Taylor (1986), Sluiter (1993: 131), Schenkeveld (1994: 270), Blank (1998: 174) and Matthaios (1999: 492). See also section 5 of this article.

[3] References to the works of Dionysius of Halicarnassus are to the chapter, page and line of the Usener / Radermacher edition (1899-1904). I have used the translations of Usher (1974-1985) for Dionysius, Householder (1981) for Apollonius Dyscolus, and Russell (2001) for Quintilian, all of which, however, I have adapted.

Law, 2003: 59).[4] Brief as it, this passage may be considered one of the very first histories of linguistics, which would make Dionysius one of the first *historians* of linguistics.[5] In this paper, I will first discuss the relationship between Dionysius's history of the word class system and the rest of his work *On Composition*. Second, I will comment on some particularities of Dionysius's 'history of linguistics'. Finally, I will compare Dionysius's approach with that of Quintilian and of modern historians of linguistics. Thus, I hope to answer the question of what kind of historian of linguistics Dionysius actually was.

2. Dionysius's history of the theory of the 'parts of speech'

Dionysius's history of the theory of the parts of speech can be found immediately after his definition of σύνθεσις (*synthesis*; composition) in the second chapter of *De compositione verborum*:

> Dionysius of Halicarnassus, *Comp.* 2.6,17-7,21:
> Ἡ σύνθεσίς ἐστι μέν, ὥσπερ καὶ αὐτὸ δηλοῖ τοὔνομα, ποιά τις θέσις
> παρ' ἄλληλα τῶν τοῦ λόγου μορίων, ἃ δὴ καὶ στοιχεῖά τινες τῆς
> λέξεως καλοῦσιν. ταῦτα δὲ Θεοδέκτης μὲν καὶ Ἀριστοτέλης καὶ οἱ
> κατ' ἐκείνους φιλοσοφήσαντες τοὺς χρόνους ἄχρι τριῶν προήγαγον, ὀνόμα
> τα καὶ ῥήματα καὶ συνδέσμους πρῶτα μέρη τῆς λέξεως
> ποιοῦντες. οἱ δὲ μετὰ τούτους γενόμενοι, καὶ μάλιστα οἱ τῆς Στωικῆς αἱρέσε
> ως ἡγεμόνες, ἕως τεττάρων προὔβιβασαν, χωρίσαντες ἀπὸ
> τῶν συνδέσμων τὰ ἄρθρα. εἶθ' οἱ μεταγενέστεροι τὰ προσηγορικὰ
> διελόντες ἀπὸ τῶν ὀνοματικῶν πέντε ἀπεφήναντο τὰ πρῶτα μέρη.
> ἕτεροι δὲ καὶ τὰς ἀντονομασίας ἀποζεύξαντες ἀπὸ τῶν ὀνομάτων
> ἕκτον στοιχεῖον τοῦτ' ἐποίησαν. οἱ δὲ καὶ τὰ ἐπιρρήματα διελόντες
> ἀπὸ τῶν ῥημάτων καὶ τὰς προθέσεις ἀπὸ τῶν συνδέσμων καὶ τὰς
> μετοχὰς ἀπὸ τῶν προσηγορικῶν, οἱ δὲ καὶ ἄλλας τινὰς
> προσαγαγόντες τομὰς πολλὰ τὰ πρῶτα μόρια τῆς λέξεως ἐποίησαν· ὑπὲρ
> ὧν οὐ μικρὸς ἂν εἴη λόγος. πλὴν ἥ γε τῶν πρώτων εἴτε τριῶν ἢ τεττάρων
> εἴθ' ὅσων δή ποτε ὄντων μερῶν πλοκὴ καὶ παράθεσις
> τὰ λεγόμενα ποιεῖ κῶλα, ἔπειθ' ἡ τούτων ἁρμονία τὰς καλουμένας
> συμπληροῖ περιόδους, αὗται δὲ τὸν σύμπαντα τελειοῦσι λόγον.

> Composition is, as the name itself indicates, a certain arrangement of the parts of speech, or the elements of diction, as some call them. Theodectes and Aristotle and the philosophers of their time increased the number of these to three, making ὀνόματα (*onomata*; nouns), ῥήματα (*rhêmata*; verbs) and σύνδεσμοι (*syndesmoi*; conjunctions) the primary parts of speech. Their

---

[4] The history of word class theory is discussed by Dionysius (*Comp.* 2.6,17-7,21) and Quintilian (*Inst. Orat.* I.4.17-21). Somewhat diverging accounts are Sch. D. Thrax, *Grammatici Graeci* I 3, pp. 515,19-521,37 (see also *Grammatici Graeci* II 3, pp. 31,26-36,24), and Priscian, *Inst.* II.15-17 (*Grammatici Latini* II, pp. 54,5-55,3).
[5] It is however possible that, for the history of the word class system, Dionysius (and Quintilian) used an older source, which is now lost. We might think of Asclepiades of Myrlea (see note 19).

successors, and in particular the leaders of the Stoic school, raised the number further to four, separating the ἄρθρα (*arthra*; articles) from the σύνδεσμοι (*syndesmoi*; conjunctions). Next, later generations distinguished the προσηγορικά (*prosêgorika*; appellative nouns) from the ὀνοματικά (*onomatika*; proper nouns) and presented the primary parts as five. Others detached the ἀντονομασίαι (*antonomasiai*; pronouns) from the ὀνόματα (*onomata*; proper nouns) and made this the sixth element. Yet others divided the ἐπιρρήματα (*epirrhêmata*; adverbs) from the ῥήματα (*rhêmata*; verbs), the προθέσεις (*protheseis*; prepositions) from the σύνδεσμοι (*syndesmoi*; conjunctions) and the μετοχαί (*metochai*; participles) from the προσηγορικά (*prosêgorika*; appellatives); while others introduced still further divisions and so made the primary parts of speech many in number. The subject could be discussed at considerable length, but it is enough to say that the combination or juxtaposition of these primary parts, whether there be three, four or any number of them, forms what are called clauses. Next, the joining together of these clauses constitutes what are called the 'periods', and these make up the complete discourse (λόγος; *logos*).

Before we take a closer look at Dionysius's history of the word class system itself, we should consider the relationship between this passage and his theory of composition. Dionysius's reason for giving a history of the word class theory is that he regards the μόρια (*moria*) or μέρη τοῦ λόγου (*merê tou logou*) as the central units of composition.[6] *Synthesis* is defined as 'a certain arrangement of the parts of speech,' and Dionysius adds that some people call these 'the elements of diction' (στοιχεῖα τῆς λέξεως; *stoicheia tês lexeôs*).[7] The parts of speech constitute the *logos*,

---

[6] Dionysius does not distinguish between μέρη and μόρια (τοῦ λόγου). The distinction between 'particles' and 'word classes' is one of later times, as has been shown by Schenkeveld (1988). In another publication, Schenkeveld (1983: 70) points out that, in Dionysius's works, ὄνομα may indicate 'proper noun' (as distinct from 'appellative'), but more often has the general sense of 'word'. In this case, Schenkeveld says, ὄνομα is used in the same way as λέξις and μέρος λόγου. I would like to add that, although ὄνομα, λέξις and μέρος λόγου can all refer to a 'word', these terms do not have the same connotations. Ὄνομα is Dionysius's most general term for 'word'. Λέξις seems to be a word qua concrete and specific form. Unlike the grammarians, he uses λέξις less frequently in the sense of 'word', which may be explained by the fact that in rhetoric and literary criticism λέξις refers to 'style', 'diction', 'expression' or 'passage'. In the case that Dionysius refers to a 'word' with the term μέρος or μόριον τοῦ λόγου, he regards it as a word that is a part or component of a larger structure. This perspective is, of course, particularly relevant in *De compositione verborum*.

[7] This is an interesting remark, for we know that the Stoic philosophers considered the parts of speech elements (στοιχεῖα). However, they did not refer to them as στοιχεῖα τῆς λέξεως, but as στοιχεῖα τοῦ λόγου: for them, the elements of λόγος were the parts of speech, while the elements of λέξις were the letters. The same distinction can be found in the works of the grammarian Apollonius Dyscolus. As far as we know, Dionysius of Halicarnassus is the only author who refers to the 'parts of speech' as στοιχεῖα τῆς λέξεως (instead of τοῦ λόγου). If it is true that the parts of speech were called στοιχεῖα τῆς λέξεως by some people, as Dionysius claims, then we might suppose that these people (or perhaps Dionysius himself?) combined an originally philosophical idea (the parts of speech as elements) with a rhetorical approach to language as expression (λέξις). However, as we have seen, Dionysius does not distinguish between μέρη λόγου and μέρη λέξεως. We should therefore not attach too much importance to his use of στοιχεῖα τῆς λέξεως instead of στοιχεῖα τοῦ λόγου.

just as the letters are the building blocks of the syllables. This Stoic view of language (see *FDS* 539-541; cf. Pinborg, 1975: 97-98 and Sluiter, 1990: 43-44) as a hierarchically structured building characterises Dionysius's entire treatment of composition:[8] *synthesis* plays a role on all levels of language, and the units on one level are the building blocks (or elements) of the units on the next level. Thus, syllables are composed of letters, words (or parts of speech) of syllables, 'clauses' of words, 'periods' of 'clauses', and the discourse of 'periods'. We may compare Apollonius Dyscolus's approach to syntax (σύνταξις), which seems to reflect the same Stoic tradition (see Apollonius Dyscolus, *Synt.* I.2 and cf. Blank, 1982: 30-31 and Sluiter, 1990: 44-46). The emphasis on the μέρη τοῦ λόγου (*merê tou logou*) in Dionysius's definition of σύνθεσις (*synthesis*) can be explained as follows: on the one hand it makes clear that, in Dionysius's view, words are the central units in the process of composition; on the other hand, it underlines the fact that words are components (*merê* or *moria*) and building blocks (*stoicheia*) of larger structures (namely clauses, periods and discourse).

When we compare Dionysius's version of the history of the word class theory with other (ancient and modern) versions, we can detect a number of interesting differences:

(1) Dionysius starts his overview with Aristotle and his student Theodectes, thereby omitting Plato, while modern historians of grammar usually observe that Plato already distinguished ὄνομα (*onoma*) and ῥῆμα (*rhêma*).[9] It is interesting, though, that Dionysius states that Theodectes and Aristotle 'increased' the number of the parts of speech: προήγαγον (*proêgagon*), the word he uses, literally means 'carried forward'. This word already contains the idea of gradual progress, which characterises the whole passage on the history of the word class system.

(2) Dionysius gives the Stoics credit for the distinction of the ὑρθρον. He attributes the distinction of the προσηγορικόν (*prosêgorikon*; appellative) to οἱ μεταγενέστεροι (*hoi metagenesteroi*), 'later people'. Since we know that the distinction between proper noun and appellative noun was definitely an invention of the Stoic philosophers, which was also known in antiquity, we might interpret the words οἱ μεταγενέστεροι as 'later generations of Stoic philosophers' (cf. *FDS* 536). I would prefer this interpretation to that of Stephen Usher, who translates οἱ μεταγενέστεροι as '[s]ubsequent *grammarians*' (Usher, 1985: 21; my italics).[10]

---

[8] The atomistic view of language is found in many other ancient texts: see also Apollonius Dyscolus, *Synt.* I.2 and IV.16.
[9] Ancient histories of the word class system never start with Plato: Quintilian (I.4.17-20) begins, like Dionysius, with Aristotle and Theodectes. See also *FDS* 543-546, overviews that start with either Aristotle or the Stoics. Modern histories that start with Plato's distinction of ὄνομα and ῥῆμα are Pinborg (1975), Robins (1966), Robins (1986), Lallot (1988) and Robins (1997⁴).
[10] More appropriate translations are those of Rhys Roberts (1910), 'later inquirers', and Aujac & Lebel (1981), 'les générations postérieures'. According to other sources, the Stoics were also responsible for the distinction of the adverb, which Antipater would have given the name μεσότης (Diogenes Laertius VII.57 = *FDS* 536). Matthaios (1999), however, has pointed out that Aristarchus (217-145 BC), who was active before Antipater of Tarsos (*fl.* 150), already knew the eight canonical word classes, including the μεσότης. Matthaios also discusses (1999: 548 ff.) the relation between Aristarchus and Antipater, and concludes that Aristarchus, like Antipater, borrowed the *term* μεσότης from older Stoic

(3) Another peculiarity is the fact that, according to Dionysius, the pronoun (ἀντονομασία; *antonomasia*) was separated from the proper noun (ὄνομα; *onoma*), whereas most ancient and modern scholars think that the pronouns, before they were recognised as a separate group, belonged to the ἄρθρα (*arthra*) (see *FDS* 542; Lallot, 1988: 17; and Robins, 1997[4]: 41).[11] The question of why Dionysius thinks that the pronoun was separated from the ὄνομα (and not from the ἄρθρον), can probably be answered by referring to the ancient grammatical theory on the ἀντωνυμία (*antônumia*; the pronoun, which Atticists like Dionysius called ἀντονομασία, *antonomasia*). According to Apollonius Dyscolus (*Synt.* I.15), the pronoun can replace the noun. Therefore, it can be combined with a verb, thus forming a complete sentence, which normally consists of a noun and a verb. Apollonius also tells us that the function of the pronoun is expressed in its name: an ἀντωνυμία, or (as Dionysius calls it) ἀντονομασία, is a word that is used 'instead of' (ἀντί; *anti*) the ὄνομα (noun).[12] Taking this theory into account, we can explain why Dionysius thinks that the pronouns were separated from the nouns (and not, as modern scholars think, from the ἄρθρον). Dionysius's idea is probably that words such as ἐγώ ('I') and σύ ('you') were originally classified as nouns (ὀνόματα), because they replaced nouns in the construction of a sentence. In later times ἐγώ and σύ would have gotten the name ἀντονομασίαι (or ἀντωνυμίαι), that is 'instead of-nouns'.

(4) A final difference between Dionysius and other historians of grammar concerns the view that the participle (μετοχή; *metochê*) was separated from the appellative (προσηγορικόν; *prosêgorikon*). According to most scholars, the participles originally belonged to the verbs (ῥήματα; *rhêmata*) before they were treated as a separate group.[13] In order to explain Dionysius's different opinion, it is again useful to take into account the ancient grammatical theory on this part of speech. The participle (μετοχή) owed its name to the fact that it 'participated' in the morphological and syntactical qualities of two other word classes, namely verb *and* noun. Apollonius

---

sources, which did, however, not give that name to a separate 'part of speech'. As far as we know, the term ἐπίρρημα (in the sense of adverb) first occurs in the fragments of Tryphon.

[11] Steinthal (II, 1891[2]: 214 ff.) follows Dionysius's view that the pronoun was separated from the noun. Matthaios (1999: 491 ff.) disagrees with Dionysius and Quintilian, but also with the traditional view that the ἀντωνυμία was separated from the Stoic ἄρθρον: the ἄρθρον, he argues, had an entirely different function than that of being a combination of two grammatical word classes, 'pronoun plus article'.

[12] The pronoun not only replaces the noun but was, according to Apollonius Dyscolus (*Synt.* I.19), even invented for the sake of the construction of verbs in the first and second person. For nouns always refer to third persons, and because verbs are also used in the first and second person, the pronoun was 'invented'. Although Apollonius Dyscolus discusses the invention of the pronouns themselves and not the invention of the term ἀντωνυμία, it is probable that Dionysius's idea on the separation of the word class 'pronoun' from the word class 'noun' is based on the same theory.

[13] See *FDS* 542:
τρίτον ὑφ' ἓν ῥῆμα ⟨καὶ⟩ μετοχή, τὸ μὲν ῥῆμα κατηγόρημα λέγοντες, τὴν δὲ μετοχὴν ἔγκλιμα ῥήματος, ὅ ἐστι ῥήματος παραγωγή. Cf. Robins (1997[4]: 41). Because of a remark by Priscian (XI.1, *Grammatici Latini* II, p. 548,1-7 [*FDS* 575]), historians of linguistics used to think that Tryphon was the first who distinguished the participle as a separate word class. However, Matthaios (1999: 420ff.) shows that Aristarchus already recognised the participle as a separate word class, for which he also used the term μετοχή.

Dyscolus explains in his *Syntax* (I.21) that participles were invented because users of language needed verbs with cases and genders, so that they could express congruence (καταλληλότης; *katallêlotês*). Thus, the participle is derived from a verb but, like a noun, it has case, number and gender. When we take into account that in ancient grammar the participle was considered a sort of intermediate form between noun and verb, it should not surprise us that Dionysius suggests that the participle was separated from the appellative, and not from the verb. We should keep in mind that the words that we call adjectives also belonged to the appellatives. It is possible that Dionysius is mainly thinking of participles that are attributively (or substantively) used.[14]

        We may conclude that, in his reconstruction of the development of the theory of the parts of speech, Dionysius is always reasoning on the basis of the name and function of the word classes that are distinguished in the system of his own time. Thus, he presumes that the pronouns originally belonged to the nouns, and that the participles were originally part of the appellatives, before these parts of speech were recognised as separate groups. Similarly, he thinks that adverbs (ἐπίρρηματα; *epirrhêmata*) originally belonged to the verbs; according to other sources, the adverbs originally belonged to the nouns. But again, we can understand that Dionysius relates the ἐπίρρημα (*epirrhêma*) to the ῥῆμα (*rhêma*). He may have thought that adverbs were considered parts of verbs, rather than that adverbs were called verbs: εὖ ποιεῖν ('to do well') would have been taken as one verb, rather than as adverb plus verb.

3. Dionysius's approach to the history of linguistics

Dionysius of Halicarnassus was, of course, not a historian of linguistics in the strict sense. As we have seen, he only mentioned the development of the doctrine of the parts of speech in the context of his own discussion of composition. Nevertheless, we might very well regard Dionysius as the first representative of a typical approach to the history of linguistics, which indeed remained the standard until the last part of the twentieth century AD.

        Generally speaking, we could distinguish between two approaches to the history of linguistics, the 'internal' and the 'external' approach (Sluiter, 1998: 24-25; see also Rorty, 1984).[15] A historian who adopts the 'internal' approach considers

---

[14] It is also possible that Dionysius relates the participles to the appellatives because both are in a way 'predicating'. In this sense, both προσηγορικά and μετοχαί are like verbs, without actually being verbs.

[15] Rorty (1984) distinguishes four genres of the historiography of philosophy: rational reconstruction; historical reconstruction; *Geistesgeschichte*; and doxography. Dionysius's 'internal' approach to the history of linguistics is related to two of Rorty's genres, namely rational reconstruction and doxography. Dionysius's historiography of linguistics belongs to the genre of rational reconstruction in the sense that he lists the views of earlier thinkers in order to solve his own problems. He reconstructs the answers that earlier thinkers would have given to his questions concerning the elements of composition, without recognizing that Aristotle and the Stoics did not have the same interests. Dionysius's account may also be described as doxography: 'the attempt to impose a problematic on a canon drawn up without reference to that problematic, or, conversely, to impose a canon on a problematic constructed without reference to that canon' (Rorty, 1984: 62). The method

earlier 'linguists' as colleagues. When dealing with a certain problem, the historian looks for solutions that have been suggested in earlier periods in the history of linguistics. He or she analyses and criticises these solutions, but does not always pay attention to the fact that earlier linguists did not ask the same questions. An ancient example of this approach is the way in which Aristotle discussed the philosophers who lived before him. As Guthrie (1957) has pointed out, Aristotle looked at the early philosophers 'in the light of his own view of reality, and [...] saw them as "striving" to reach the same view' (Guthrie, 1957: 38). The second approach to the history of linguistics is called the 'external' approach. The historian who adopts this method does not try to apply earlier linguistic theories to his or her own purpose. Instead, he or she attempts to take into account the context in which earlier ideas about language were developed, and adheres to the 'principle of charity' (Sluiter, 1998: 25).

It is clear that Dionysius of Halicarnassus belongs to the group of historians who adopt the 'internal' approach to the history of linguistics. He discusses the history of the word class system only because he has to find an answer to the question as to which elements are the central units one uses when composing sentences and texts. Aristotle, the Stoics and the grammarians were, of course, dealing with different problems, but Dionysius applies their views, which originated in such diverse fields as ontology, logic, philology or grammar, to the topic of his own investigation into σύνθεσις (synthesis).

The internal method in the historiography of science, as we find it in Aristotle and Dionysius, is often combined with a strong belief in progress. The traditional historian of linguistics looks back from the standpoint of his own linguistic system and considers earlier periods as preliminary stages that were already groping for and striving towards that system (cf. Schmitter, 1987: 103). This attitude is particularly characteristic of nineteenth century scholars such as Benfey (1869: 121 ff.) and Steinthal (II, 1891²: 209-218) (cf. Grotsch, 1982: 118-139 and Schmitter, 1987: 105). But even a more recent scholar like Robins, in spite of his own warnings against the dangers of 'looking to the past through the eyes of the present', presents the development of the word class theory in a tree diagram (Robins, 1986: 20), which bears a remarkable resemblance to the scheme that one can extract from the second chapter of Dionysius of Halicarnassus's *On Composition*.[16] Robins presents the system of eight word classes as the result of a long cumulative process: Plato, Aristotle, the Stoics and the grammarians, it is suggested, all contributed their bit to the completion of the final word class system. As we have seen, the idea of progress is also clearly present in Dionysius's account: 'Theodectes and Aristotle *increased* the number of the

---

that Sluiter calls the 'external' approach corresponds to Rorty's historical reconstruction. It concentrates on the historical contexts in which early ideas (on philosophy, or language, for that matter) were developed.

[16] For a theoretical discussion of the problematic notion of 'progress', see Schmitter (1987: 103-113). Robins (1997⁴: 3) warns against a presentation of the history of linguistics in terms of progress: 'It is tempting, and flattering to one's contemporaries, to see the history of a science as the progressive discovery of the truth and the attainment of the right methods [...]. But this is a fallacy.' Nevertheless, the latest edition of his *A Short History of Linguistics* (1997⁴: 31-46) still presents the ancient history of the theory of the parts of speech in the traditional way. Robins's other surveys (1966 and 1986) are similar in this respect. Possibly, Dionysius of Halicarnassus and Quintilian were his primary sources. For an analysis of Robins's approach to the history of linguistics, see Grotsch (1982: 147-150).

parts of speech to three; the Stoics *raised* the number further to four; [...] others made the primary parts of speech many in number.' In fact, the resemblance between Robins and Dionysius is of course not so remarkable at all. By now it has become clear that the traditional approach to the history of linguistics, which tends to portray the history of linguistic ideas as the 'progressive discovery of the truth' (Robins, 1997[4]: 3), can be largely traced back to Dionysius's *On Composition*.

There is, however, one important aspect in which Dionysius differs from later historians of linguistics. Unlike later scholars, Dionysius does not present the history of the word class theory as leading to a final and complete system of eight or nine μέρη τοῦ λόγου (*merê tou logou*). Although he implicitly mentions a system of nine, he adds that other people distinguished even more parts of speech. Dionysius does not express his preference for a particular system, and in the end he does not seem to care how many parts of speech really exist. It is unimportant to him 'whether there be three, four or any number of them.' This attitude is reflected in other parts of his work, where Dionysius leaves open the question of how certain words should be classified. He tells us, for instance, that the word ἐπί ('on') might be called either a σύνδεσμος (*syndesmos*; connective) or a πρόθεσις (*prothesis*; preposition) (*Comp.* 22.102,16).[17] Such remarks do not only indicate that in Dionysius's time the system of eight word classes had not yet become a fixed canon, but also that the exact number of word classes was not so important for Dionysius's specific purpose. For the composition of a text out of words, it does not matter to which particular word classes these words belong. For a historian of linguistics who was more inclined to view the word class system of his own time as the final truth about the matter, we have to turn to Quintilian, whose *Institutio oratoria* was written at the end of the first century AD.

4. Quintilian's history of the theory of the parts of speech

The similarities between the passages of Dionysius (*Comp.* 2) and Quintilian (*Inst. Orat.* I.4.17-21) have often been noted (Colson, 1924: 45-46, Schenkeveld, 1994: 270 n. 22 and Matthaios, 1999: 194 n. 17). Quintilian's account of the development of the word class theory is as follows (on this passage, see also Colson, 1914, 1916, 1924):

> Quintilian, *Inst. Orat.* I.4.17-21:
> *Tum uidebit, ad quem hoc pertinet, quot et quae partes orationis, quamquam de numero parum conuenit. Veteres enim, quorum fuerunt Aristoteles quoque atque Theodectes, uerba modo et nomina et coniunctiones tradiderunt, uidelicet quod in uerbis uim sermonis, in nominibus materiam (quia alterum est quod loquimur, alterum de quo loquimur), in coniunctionibus autem complexum*

---

[17] One might think that this remark only points to the difficulty of the *merismos* procedure, and that it merely makes clear that Dionysius thinks that the word ἐπί could be assigned to the prepositions as well as to the conjunctions. However, Schenkeveld (1983: 73-74) makes it clear that the alternative classifications are related to the existence of different word class systems: the classification of ἐπί as a *syndesmos* 'is a sure sign of a system with less than nine (or eight) parts.' In a system with fewer parts of speech, the *syndesmos* would also have covered words that were in later times identified as prepositions (which the Stoics called *syndesmoi prothetikoi*).

eorum esse iudicauerunt: quas coniunctiones a plerisque dici scio, sed haec uidetur ex syndesmo magis propria tralatio. Paulatim a philosophis ac maxime Stoicis auctus est numerus, ac primum conuinctionibus articuli adiecti, post praepositiones: nominibus appellatio, deinde pronomen, deinde mixtum uerbo participium, ipsis uerbis aduerbia. Noster sermo articulos non desiderat ideoque in alias partes orationis sparguntur, sed accedit superioribus interiectio. Alii tamen ex idoneis dumtaxat auctoribus octo partes secuti sunt, ut Aristarchus et aetate nostra Palaemon, qui uocabulum siue appellationem nomini subiecerunt tamquam speciem eius, at ii qui aliud nomen, aliud uocabulum faciunt, nouem. Nihilominus fuerunt qui ipsum adhuc uocabulum ab appellatione diducerent, ut esset uocabulum corpus uisu tactuque manifestum: 'domus' 'lectus', appellatio cui uel alterum deesset uel utrumque: 'uentus' 'caelum' 'deus' 'uirtus'. Adiciebant et adseuerationem, ut 'eheu', et tractionem, ut 'fasciatim': quae mihi non adprobantur. Vocabulum an appellatio dicenda sit προσηγορία et subicienda nomini necne, quia parui refert, liberum opinaturis relinquo.

The teacher responsible will then need to consider how many parts of speech there are, and what they are, although there is little agreement about the number. Earlier writers, including also Aristotle and Theodectes, listed only verbs (*verba*), nouns (*nomina*), and 'convinctions' (*convinctiones*). Evidently, they took the active element in language to be in the verbs, and the material element in the nouns, because the one is what we say, the other is what we speak about, while the 'convinctions' provided the connections between them. I know most people say 'conjunctions', but 'convinctions' seems the better translation of *syndesmos*. The philosophers, particularly the Stoics, gradually increased the number: to 'convinctions' were first added 'articles' (*articuli*), and then 'prepositions' (*praepositiones*); to 'nouns' was added the 'appellation' (*appellatio*), next the 'pronoun' (*pronomen*), and then the quasi-verbal 'participle' (*participium*); to 'verbs' were added 'adverbs' (*adverbia*). Our language does not need articles (*articuli*), and these are therefore distributed among other parts of speech. In addition, however, there is the interjection (*interiectio*). Some, belonging to the competent authorities, have gone as far as eight parts of speech:[18] so Aristarchus and, in our own day, Palaemon, who both put 'vocable' or 'appellative' under 'noun', as species of that genus. Those who distinguished 'vocable' from 'noun' make the total nine. Yet some have also separated 'vocable' itself from 'appellation', making 'vocable' indicate visible and tangible objects — 'house' or 'bed' — and 'appellation' things in which either or both of these characteristics were absent, like 'wind', 'heaven', 'God', or 'virtue'. They have also added 'asseveration' (like *eu*) and 'derivative' (like *fasciatim*). I do not approve of these. Whether we should

---

[18] Most translators take the words *ex idoneis auctoribus* with *secuti sunt*: 'others followed good authorities.' Russell (2001) translates 'some, with good authorities to back them.' It is, however, more probable that Quintilian considered Aristarchus and Palaemon themselves as 'competent authorities'. Therefore, I agree with Matthaios (1999: 191 n. 2), who interprets *ex idoneis auctoribus* as a partitive construction. For the expression *idonei auctores*, see also Kasten (1978).

translate *prosêgoria* as 'vocable' or as 'appellation', and whether it should be regarded as a subclass of the noun, is an unimportant question, and I leave it open to personal opinions.

In this paper, I cannot draw a systematic comparison of Dionysius's and Quintilian's versions of the history of the word class system. There are many similarities between the two accounts, and it is probable that Quintilian made use of Dionysius, or that the two versions are based on the same source.[19] There are, however, some differences as well. One interesting difference is the fact that Dionysius constantly speaks of 'splitting' and 'separation', whereas Quintilian refers to the 'addition' and 'extension' of the system. Dionysius uses the words χωρίζω (*chôrizô*; 'to separate'), διαιρέω (*diaireô*; 'to divide'), ἀποζεύγνυμι (*apozeugnumi*; 'to part') and τομή (*tomê*; 'division'), while Quintilian uses the verbs *adicio* ('to add') and *accedo* ('to join', 'to be added').[20] The different vocabulary reflects a difference in perspective. Dionysius reasons from the past and emphasises the many distinctions that were developed in the course of time, while Quintilian presents the history of the word class theory as gradually leading to the completion of the system in his own time.

These diverging perspectives are related to another difference between the two accounts. While Dionysius, as we have seen, does not really care how many parts of speech exactly exist, 'whether there be three, four or any number of them,' Quintilian demands that the teacher of grammar is clear about the question how many parts of speech there are, and what they are: *quot et quae partes orationis*. These words remind us of the opening of Donatus's *Ars minor*, which I quoted at the beginning of this article. Although Quintilian admits that there is no agreement on the exact number, he clearly opts for a system of eight or nine parts of speech, and he explicitly rejects the later additions to the system (*quae mihi non adprobantur*, 'which are not approved by me'). To explain the different attitudes of Dionysius and Quintilian, we should look at the contexts in which they were presenting their histories of the word class system. In Dionysius's account, the word classes figure as the primary building blocks of composition. Certainty about the exact number of these 'elements' was not relevant for Dionysius's purpose, since, when one composes a text, it does not really matter whether one assigns a word to one word class or another. Quintilian, on the other hand, discussed the history of the word class system in a passage about the teaching of Latin and Greek in the school of the grammarian. The procedure of *merismos* (the classification of the parts of speech) was a standard exercise in the lessons of the *grammaticus*, so that clarity about the number of word classes was necessary. Obviously, from a pedagogical point of view, one would not want to bother one's

---

[19] Blank (1998: xlv-xlvi) has argued that much of the grammatical theory found in both Sextus Empiricus (2nd century AD) and Quintilian can be traced back to Asclepiades of Myrlea, who possibly taught in Rome in the early first century BC. Sextus Empiricus does not refer to the history of the word classes, but we should not rule out the possibility that Asclepiades was the source of the accounts of Dionysius and Quintilian.

[20] With respect to the *number* of 'parts of speech', both Dionysius and Quintilian speak in terms of extension. Dionysius uses the words προήγαγον, προύβιβασαν, προσαγαγόντες, while Quintilian says *auctus est*.

students too much with the different views that various scholars had developed on the subject.[21]

5. Conclusion: Dionysius, Quintilian and modern historians of linguistics

In his important article 'Rethinking the History of Language Science in Classical Antiquity' (1986), Taylor states that one of the key notions central to the traditional version of Graeco-Roman language science is 'the emphasis upon the development of the doctrine of the parts of speech, especially as it accumulates or evolves in measured stages from its beginnings in Plato to its fullest expression in Dionysius Thrax' (Taylor, 1986: 177). In this paper, I have tried to show in what sense Dionysius of Halicarnassus and Quintilian can be considered the prototypes of modern traditional historiographers of ancient linguistics.

Dionysius's history of the word class system is in two respects characteristic for the traditional historiography of linguistics. First, he adopts an 'internal' approach to the history of science, applying earlier views on language, which were developed in several disciplines, to his own particular subject, which is in his case the art of composition. Second, his account of the development of the word class theory is characterised by the idea that *gradual progress* was made by successive stages in the history of linguistics. Unlike many other historians of linguistics, however, Dionysius does not present the word class system of his own time as the ultimate truth. Quintilian, on the other hand, expresses his preference for a system with eight or nine parts of speech. I have explained this difference by pointing to the different contexts in which the two writers presented their views.[22]

---

[21] Another difference between Dionysius and Quintilian is the following: Dionysius distinguishes five stages in the development of the theory of the parts of speech, while Quintilian summarises these in only two stages, to which he adds two Roman developments of the system. The four stages in Quintilian's overview are organised in the following way: (1) like Dionysius, Quintilian starts with Aristotle and Theodectes, who would have known three parts of speech; (2) Next, Quintilian states that the number of parts of speech increased 'gradually' (*paulatim*) but, unlike Dionysius, he does not in the first instance present the extension of the system chronologically, but *systematically*. The starting point is the system of Aristotle, and the new word classes are discussed in relation to the three original ones, namely σύνδεσμος (*convinctio*), ὄνομα (*nomen*), and ῥῆμα (*verbum*). Within his presentation of the development of the system Quintilian does make chronological distinctions, by adding words like *primum* ('first'), *post* ('next') and *deinde* ('thereafter'). Quintilian's second stage includes the same word classes as Dionysius's fifth stage; (3) The third stage in Quintilian's overview is the Roman substitution of the interjection for the article. Quintilian remarks that some people put the appellative under 'noun' ('as species of that genus'), while other people consider *vocabulum* and *nomen* as two different word classes. That makes the total number of parts of speech eight or nine; (4) In a fourth stage, even more distinctions were added by 'others' (*alii*): *vocabulum, adseveratio*, and *tractio*; but Quintilian himself rejects these differentiations. The additions to the system that he mentions would increase the total number of word classes to a maximum of twelve, but Quintilian himself opts for a system of eight or nine word classes.

[22] For the later historiography of linguistics, Priscian (6th century AD) was of course very influential: he presented a history of the word class theory and adhered to a fixed number of eight *partes orationis*: see *Grammatici Latini* II, pp. 54,5-55,3. Similar is the *Ars anonyma Bernensis* (*FDS* 549). Donatus (*Grammatici Latini* IV, p. 372), who uses a system of eight parts of speech, does not discuss the history of the word class system, but only remarks that *multi plures, multi pauciores partes orationis putant.*

Over the last two decades, Taylor (1986) and other historians of ancient linguistics, such as Sluiter (1993, 1998), Schenkeveld (1994) and Law (2003), have distanced themselves from the traditional approach to the history of linguistics in general and to the history of word class theory in particular. Nowadays, scholars are more willing to recognise that Plato, the Stoics, the Alexandrian philologists and the technical grammarians all had their own different purposes and, accordingly, that the units they called τὰ μέρη τοῦ λόγου *(ta merê tou logou)* were entirely different matters for all of them.[23] In the article mentioned above, Taylor (1986) stated that the different philosophers, philologists and grammarians 'were not playing the game by the same rules' (Taylor 1986: 179). I would like to go one step further: *they were not even playing the same game.* Philosophers were not interested in enumerating as many word classes as possible, so one would do them wrong by interpreting them as if they were grammarians. As a *historian of linguistics,* therefore, I do not agree with the way in which Dionysius and Quintilian presented the history of the word class system. As a *historian of the historiography of linguistics,* however, I conclude that their approach to the history of linguistics has been very influential.[24]

Sources

Apollonius Dyscolus: Jean Lallot. 1997. *Apollonius Dyscole, De la construction (syntaxe).* Texte grec accompagné de notes critiques, introduction, traduction, notes exégétiques, index. Paris: Librairie Philosophique J. Vrin.

Dionysius of Halicarnassus: Hermann Usener & Ludwich Radermacher. 1899-1904. In: *Dionysii Halicarnasei quae exstant,* Vol. V & VI (Opuscula I & II). Stuttgart / Leipzig: Teubner.

Greek grammarians: 1867-1901. *Grammatici Graeci.* Lipsiae: Teubner (reprint 1965. Hildesheim: Olms).

Quintilian: Michael Winterbottom. 1970. *M. Fabi Quintiliani Institutionis Oratoriae libri duodecim.* Oxford: Oxford University Press.

Roman grammarians:1855-1880. *Grammatici Latini.* Lipsiae: Teubner (reprint: 1961. Hildesheim: Olms).

Stoic philosophers *(FDS)*: Karlheinz Hülser. 1987-1988. In: *Die Fragmente zur Dialektik der Stoiker, Neue Sammlung der Texte mit deutscher Übersetzung und Kommentaren [= FDS].* Stuttgart: Frommann-Holzboog.

Translations

Apollonius Dyscolus: Fred W. Householder. 1981. *The Syntax of Apollonius Dyscolus* (translation and commentary). Amsterdam: John Benjamins.

---

[23] See now also Matthaios (1999: 492): 'Die von Dionysios van Halikarnaß und Quintilian gegebene Erklärung für die Erweiterung des Wortartensystems durch Aufspaltung umfangreicher Redeteile läßt genauso wie die in den grammatische Berichten voorgenomen Zuweisung der einzelnen Wortarten zum stoischen Redeteilsystem die Tatsache außer acht, daß der Begriff "Redeteil" bzw. "Wortart" von Schultradition zu Schultradition eine andere Bedeutung hat.'

[24] Earlier versions of this paper were presented at the *Societas Linguistica vulgo Katwijk* (Katwijk 2003) and at the *Annual Colloquium of the Henry Sweet Society for the History of Linguistic Ideas* (Oxford 2004). I wish to thank Ineke Sluiter, Therese Tiedemann, Almuth Wietholtz and Jan van Ophuijsen for their useful suggestions.

Dionysius of Halicarnassus: Stephen Usher. 1974-1985. *Dionysius of Halicarnassus, The Critical Essays in Two Volumes*, with an English translation. Cambridge, Massachusetts / London, England: Harvard University Press.

Quintilian: Donald A. Russell. 2001. *Quintilian, The Orator's Education*. Cambridge, Massachusetts / London, England: Harvard University Press.

References

Aujac, Germaine & Maurice Lebel. 1981. *Denys d'Halicarnasse, Opuscules Rhétoriques, Tome III, La Composition Stylistique.* Paris: Les belles lettres.

Benfey, Theodor. 1869. *Geschichte der Sprachwissenschaft und orientalischen Philologie in Deutschland seit dem Anfange des 19. Jahrhunderts mit einem Rückblick auf die früheren Zeiten.* München: Oldenbourg.

Blank, David L. 1982. *Ancient Philosophy and Grammar. The Syntax of Apollonius Dyscolus (American Classical Studies 10).* Chico, California: Scholars Press.

-----. 1998. *Sextus Empiricus, Against the Grammarians (Adversus Mathematicos I),* Translated with an introduction and commentary. Oxford: Oxford University Press.

Colson, Francis Henry. 1914. 'The Grammatical Chapters in Quintilian I.4-8.' *Classical Quarterly*, 8: 33–47.

-----. 1916. 'Some Problems in the Grammatical Chapters of Quintilian.' *Classical Quarterly*, 10: 17–31.

-----. 1924. *M. Fabii Quintiliani Institutionis Oratoriae Liber I.* Edited with introduction and commentary. Cambridge: Cambridge University Press (reprint 1973. Hildesheim: Olms).

Grotsch, Klaus. 1982. *Sprachwissenschaftsgeschichtsschreibung. Ein Beitrag zur Kritik und zur historischen und methodologischen Selbstvergewisserung der Disziplin.* Göppingen: Kümmerle.

Guthrie, William K.C. 1957. 'Aristotle as a Historian of Philosophy: Some Preliminaries.' *Journal for Hellenic Studies*, 77: 35–41.

Kasten, Robert A. 1978. 'Servius and idonei auctores.' *American Journal of Philology*, 99: 181-209.

Lallot, Jean. 1988. 'Origines et développement de la théorie des parties du discours en Grèce.' *Langages*, 92: 11–23.

-----. 1992. 'Le mot, dans la tradition prégrammaticale et grammaticale en Grèce.' *LALIES. Actes des sessions de linguistique et de littérature*, 10: 125–134.

Law, Vivien. 2003. *The History of Linguistics in Europe from Plato to 1600.* Cambridge: Cambridge University Press.

Matthaios, Stephanos. 1999. *Untersuchungen zur Grammatik Aristarchs: Texte und Interpretation zur Wortartenlehre (Hypomnemata 126).* Göttingen: Vandenhoeck & Ruprecht.

Pinborg, Jan. 1975. 'Classical Antiquity: Greece.' In: *Historiography of Linguistics (Current Trends in Linguistics 13).* Ed. by Thomas A. Sebeok. Den Haag & Paris: Mouton. 69–126.

Rhys Roberts, W. 1910. *Dionysius of Halicarnassus On Literary Composition.* London: Macmillan.

Robins, R.H. 1966. 'The Development of the Word Class System of the European Grammatical Tradition.' *Foundations of Language*, 2: 3–19.

-----. 1967 [1997[4]]. *A Short History of Linguistics*. London & New York: Longman.

-----. 1986. 'The Technê grammatikê of Dionysius Thrax in its Historical Perspective: the Evolution of the Traditional European Word Class Systems.' In: *Mot et Parties du Discours, Word and Word Classes, Wort und Wortarten*. Ed. by P. Swiggers & W. van Hoecke. Leuven & Paris: Peeters. 9–37.

Rorty, Richard. 1984. 'The Historiography of Philosophy: Four Genres.' In: *Philosophy in History. Essays on the Historiography of Philosophy*. Ed. by Richard Rorty, J.B. Schneewind, Quentin Skinner. Cambridge: Cambridge University Press. 49–75.

Schenkeveld, Dirk M. 1983. 'Linguistic Theories in the Rhetorical Works of Dionysius of Halicarnassus.' *Glotta*, 61: 67–94.

-----. 1988. 'From Particula to Particle — The Genesis of a Class of Words.' In: *L'héritage des grammairiens latins de l'antiquité aux lumières, Actes du Colloque de Chantilly 2-4 septembre 1987*. Ed. by Irène Rosier. Paris: Peeters. 81–93.

-----. 1994. 'Scholarship and Grammar.' In: *La philologie grecque à l'époque hellénistique et romaine. (Entretiens sur l'Antiquité classique 40)*. Ed. by F. Montanari. Genève: Vandoeuvres. 263–306.

Schmitter, Peter. 1987. 'Fortschritt. Zu einer umstrittenen Interpretationskategorie in der Geschichtsschreibung der Linguistik und der Semiotik.' In: *Geschichte der Sprachtheorie 1. Zur Theorie und Methode der Geschichtsschreibung der Linguistik*. Ed. by Peter Schmitter. Tübingen: Gunter Narr Verlag. 93–124.

Sluiter, Ineke. 1990. *Ancient Grammar in Context. Contributions to the Study of Ancient Linguistic Thought*. Amsterdam: VU University Press.

-----. 1993. 'Antieke grammatica: autonoom of instrument?' In: *Nieuwe wegen in taal- en literatuurwetenschap, Handelingen van het Eenenveertigste Filologencongres*. Ed. by J. Goedegebuure. Tilburg: Tilburg University Press. 129–141.

-----. 1998. 'Metatexts and the Principle of Charity.' In: *Metahistoriography. Theoretical and Methodological Aspects of the Historiography of Linguistics*. Ed. by Peter Schmitter & Marijke van der Wal. Münster: Nodus. 11–27.

Steinthal, Heymann. 1863 [1890[2] (I) - 1891[2] (II)]. *Geschichte der Sprachwissenschaft bei den Griechen und Römern, mit besonderer Rücksicht auf die Logik*. Berlin: Dümmler (reprint Hildesheim 1971).

Taylor, Daniel J. 1986. 'Rethinking the History of Language Science in Classical Antiquity.' *Historiographia Linguistica*, 13: 175–190.

-----. 1987. 'Rethinking the History of Language Science in Classical Antiquity.' In: *The History of Linguistics in the Classical Period (Amsterdam Studies in the Theory and History of Linguistic Science 46)*. Ed. by Daniel J. Taylor. Amsterdam / Philadelphia: John Benjamins. 1–16. Reprint of Taylor (1986).

Contact details: **c.c.de.jonge@let.leidenuniv.nl**

The Notion of Sememe in the Work of Adolf Noreen

Serhii Vakulenko
Kharkiv, Ukraine

The term *sememe*, probably because of its distinct French flavour, is most readily associated with the so-called structural semantics that flourished in France in the 1960s, best represented by the works of scholars like Algirdas Greimas (1917–1992) and Bernard Pottier. In spite of the numerous conceptual affinities between the two, their terminological usage was very far from being uniform (cf. Lavorel, 1975: 63), and these divergences did concern, among others, the concept of sememe. Greimas (1966: 44f, 51) opposed the *sememe*, considered to be a phenomenon of speech, uniting a semic core and a number of contextual semes, to the *lexeme*, seen as a virtual model subsuming the whole functioning of a figure of signification covered by a certain formant, but anterior to any speech manifestation. Pottier, instead, defined it simply as a set of semes, or distinctive semantic features, that constitute the meaning of a lexical morpheme (cf.: Pottier, 1963: 15; 1964: 124; 1964, 1973: 453; Pottier, Audubert and Pais, 1973: 41, 44). It is Pottier's acceptation that prevailed in the long run, gaining considerable diffusion over large portions of the linguistic scene, including, especially, Eastern Europe, where it has become part and parcel of everyday scholarly usage and is integrated into standard university textbooks either in a form close to Pottier's original wording (cf.: Bidu-Vrănceanu and Forăscu, 1984: 17; Filipec and Čermak, 1985: 66), or with the important explicit specification that in the case of a polysemous word each separate meaning constitutes an independent sememe (cf. Oleksenko, 1999: 9).[1]

In the British Isles, on the contrary, the term in question does not seem to enjoy much popularity. Two rather telling examples might suffice to illustrate this tepidness. One would look in vain for the term *sememe* in John Lyons' *Semantics* (1977), quite a landmark in the history of the discipline. In fact, Lyons (1977: 20) prefers to speak not of *sememes*, but of the *meanings of lexemes*, and avoids using the word even in his discussion of *semes* and the structure of *lexical fields*, although he cites Pottier amply on this occasion (1977: 326). Nor is the *sememe* to be found in the popular *Cambridge Encyclopedia of Language* by David Crystal, a voluminous reference work that has enough room to include, for example, a term like *mytheme* (1987: 79).

One reason for the omission might lie in the unclear ontological status of the sememe. It is well known that structural semantics was largely patterned on the theoretical models developed several decades earlier in the domain of phonology, relying on the assumption of an isomorphism between the plane of expression and the plane of content. The search for analogies produced two basic approaches that may be associated with the names of the American scholars Ward Hunt Goodenough and Floyd Glenn Lounsbury (1914–1998). According to Rodney Huddleston (1974: 5), the

---

[1] This specification constituted, in fact, the first – practically immediate – reaction to Pottier's terminological proposals on the part of some German scholars (cf.: Heger 1964: 509–512; Meier 1964).

author of a study devoted especially to the sememe, these two approaches can be summarized in the form of two equations that show the place occupied by the sememe in the language system leaning on the much more evident (or at least better known) relationship between the basic units of linguistic expression – morpheme, phoneme, and distinctive phonetic feature:

1. *W.H. Goodenough*:
phoneme   : morpheme   = semantic   : sememe
                                          component

2. *F.G. Lounsbury*:
distinctive   : phoneme   = distinctive   : sememe
phonetic   semantic
feature   feature[2]

       In spite of a rather obvious asymmetry between the two equations, they agree, as Huddleston (1974: 5) notes, in that the undoubtedly independent status of the first three forms seems to confer an equal status on the fourth term, namely, the *sememe*. As far as Goodenough's formula is concerned, Huddleston (1974: 3ff) sees its weak point in the fact that morphemes and phonemes are distinct basic units belonging to two different levels and established with the help of distributional factors that have no parallel in semantics. It is not altogether clear, though, why Huddleston is so categorical about Goodenough's adhesion to this view of isomorphism. In his treatment of the Trukese kinship terms, Goodenough (1956: 200ff), in fact, uses the example of the lexemes *semej* and *jinej* that share the semantic components (A) *tefej* (signifying "the universe of kinship" in general), and (B$_1$) "of higher generation", but differ with respect to the third variable "sex", having, accordingly, the values C$_1$ ("male") and C$_2$ ("female"), and then proceeds to state the structural analogy of such concept forms to morphemes (1956: 208): "The structural phonological relationship of English *pet* and *bet* [...] parallels the structural conceptual relationship of AB$_1$C$_1$ (*semej*) and AB$_1$C$_2$ (*jinej*)." At the same time, though, his article (1956: 197) contains yet another comparison:

          [...] sememes consist of combinations of percepts and/or concepts, which
          consist of values for different variables. Just as we may say that the English
          phonemes /t/ and /d/ complement each other with respect to the variable of
          voicing, or /p/, /t/, and /k/ are a complementary set with respect to place of
          articulation, it is possible for us to arrange sememes into complementary sets,
          e.g. the sememes of *old* and *young* complement each other with respect to a
          conceptual variable of age, *come* and *came* with respect to tense, *husband* and
          wife with respect to sex.

---

[2] The terms *semantic component* and *distinctive semantic feature* can be regarded as synonymous, and, moreover, as equivalent to *seme* in the European structuralist tradition, although both Lounsbury (1956: 192) and Goodenough (1956: 208) use the term *seme* in an acceptation different from Pottier's.

This second analogy, found in Goodenough, is undoubtedly identical with the one Huddleston (1974: 4) attributes to Lounsbury, which holds "between the phoneme as a bundle of distinctive phonetic features and the sememe as a bundle of distinctive semantic features." Since the terms on the left-hand side of the second equation belong to the same level, the second equation is free from the flaw inherent in the first one. Nevertheless, it is marred by the lack of any analogue in semantics to the linearity of the phonological level. Accordingly, the arrangement of distinctive phonetic features within a phoneme is of a different kind from that of phonemes relative to each other, whereas semantics displays no systematic difference between the arrangement of components within sememes and that of sememes relative to each other (cf. Huddleston, 1974: 4f).

A certain lack of philological accuracy in Huddleston's analysis does not impair the substance of his (1974: 6) theoretical conclusion concerning both equations, according to which "the sememe is not an independent unit of semantic structure" and so is altogether dispensable within the terminological outfit of structural semantics. It could be noted in passing that it had in fact no function in one of the most influential currents of structural semantics, Eugen Coşeriu's (1921–2002) *lexematics*, with its many followers in Germany and elsewhere, albeit Coşeriu himself was ready to acknowledge, on the whole, Pottier's merits in establishing "a linguistic apparatus which is most highly suitable for application to the content-analysis of lexical fields" (cf. Coseriu and Geckeler, 1974: 135).

For the present story, however, it is more important that both Lounsbury (1956) and Goodenough (1956) published their pioneering articles on componential analysis of lexical meanings several years ahead of the development of structural semantics in France. Still, as far as their use of the English term *sememe* is concerned, they were no innovators. Indeed, the English *sememe* antecedes its French counterpart *sémème* by several decades, having been introduced into common usage by Leonard Bloomfield (*Language* [1933]).[3] It is true that although Bloomfield (1950: 74) saw linguistics, "on an ideal plane," as consisting of "two main investigations," *phonetics* and *semantics*, he did not attempt to establish any kind of isomorphism between their respective objects. As far as the lexical level of linguistic signalling is concerned, his theory (1950: 161f, 166, 264) states that *phonemes*, the smallest *meaningless* signalling-units, enter in different combinations to form *morphemes*, the smallest *meaningful* units, whose meanings are called, precisely, *sememes*. Of course, Bloomfield only educed the notion of sememe the better to shut it out from the scope of linguistics, inasmuch as he was convinced that a linguist can hardly say anything sensible about the meanings of speech-forms, given the fundamental imperfection of our knowledge of the world we live in (1950: 74, 162, 276). Accordingly, the linguist can merely assume that each sememe is "a constant and definite unit of meaning, different from all other

---

[3] The second edition of the *Oxford English Dictionary* (but not the first one) lists some attestations of the word in question prior to 1933, most of them clearly inspired by Scandinavian sources and reducible to Noreen's terminological innovation that will be discussed in detail later on. Bloomfield's use of the word (as early as 1926) seems, on the contrary, quite spontaneous. His priority in this respect was always readily recognized by the American scholars. This applies, among others, to Lounsbury (1956: 192, fn. 49) and Goodenough (1956: 208), even though both rejected Bloomfield's explanation of the notion in question.

meanings" and arbitrarily conjoined to a certain combination of phonemes, but not analyzable with the methods used in his science (1950: 162).

In German, the first attestation of the term in question dates from 1923. In the form *Semem*, it appears in Adolf Gotthard Noreen's (1854–1925) book *Einführung in die wissenschaftliche Betrachtung der Sprache: Beiträge zur Methode und Terminologie der Grammatik*. However, the Germans have no right to claim any sort of priority in this case, not only because Noreen was a Swede, but also because the book in question (Noreen, 1923) was, as its subtitle specifies, "an authorized and revised translation of selected parts of the Swedish work *Vårt språk* [*Our Language*]." As such, it actually contains nothing new in comparison with the original text, although it has the merit of summarizing within its 460 pages Noreen's main theoretical viewpoints dispersed all over the mammoth multivolume Swedish edition. Of particular moment for our topic is the stress it lays – mirrored in its very title – on the terminological aspects of grammar. Indeed, Noreen's concern with the inadequacy of current terminological usage accompanied him throughout his scholarly career. As early as 1879, at the age of 25, he published an important article "Något om ord och ordklasser" ("Some Remarks on Words and Word-Classes"), in which he deplored the fact that "stable and precise terminology"[4] is altogether lacking in the field of the science of language (1879: 18). The continuous efforts on his part to improve that unfortunate situation did make his own terminology "extremely deliberate" (in Swedish, *ytterst genomtänkt*), to repeat Benny Brodda's (1973: 5) expression. In particular, he can be considered, according to Aarni Penttilä (1971: 9), the actual initiator of the so-called "emic" linguistics, as he was the first scholar to establish some of its most basic terms, like *morpheme, phoneme*, and *sememe*. However, Noreen (1904: 6) himself was more modest and only claimed to have coined the term *sememe*, while admitting that the borrowed French terms *phoneme* and *morpheme* served as his source of inspiration. In fact, the 5th volume of *Vårt språk*, printed in separate fascicles from 1904 to 1912 – and thus sharing the complicated publishing history of the work as a whole (which ultimately remained unfinished) (cf. Lotz, 1954) – was not the place were the Swedish term *semem* first appeared. It had been introduced in Noreen's booklet entitled *Inledning till modermålets betydelselära* [*Introduction to the Semantics of the Mother Tongue*], which contained a summary of a cycle of six lectures he had held at the 1901 Uppsala Summer Course for schoolteachers (cf. Löwkrantz, 1901).[5] Thus, the year 1901 may be reasonably safely taken as the birth-date of the term in question (with a prolonged period of "pregnancy" necessarily preceding it).

Actually, this terminological innovation springs from Noreen's tripartite division of the science of language, already present in his 1879 article on words and word-classes. Noreen (1879: 21) explicitly states there that any linguistic expression can be considered from two viewpoints: that of its material side and that of its content. The material side, represented by the various articulate sounds, is the subject of the discipline he calls *ljudlära* in Swedish (literally, the science of sounds), while the

---

[4] Original: "[…] fast och precis terminologi […]"
[5] Symptomatically, Löwkrantz does not use the term *semem* (which must have seemed rather obstruse to him) in his popular account of the content of the lectures, although he mentions some of Noreen's basic ideas connected with the notion of sememe.

content side falls asunder, once more, into two parts, and can be approached either from the point of view of *construction*, or *form*, or from the point of view of its *function*, or *meaning*. The proposed names of the respective disciplines, in Swedish, are *konstruktionslära* (or − preferably − *formlära*), and *betydelselära*. However, this distinction suffers a slight transformation in the first volume of *Vårt språk*, where Noreen (1903: 50) speaks not of two, but of three fundamental and equipollent viewpoints:

> As language is essentially − like clothes, dwellings and tools − an artificial product, it must offer for examination just as many and just the same principal aspects as any other artificial product, namely that of m a t e r i a l (the stuff of which it was made), that of c o n t e n t   (the substance it "represents" or "handles"; the task it has to "fulfil", its goal), and that of f o r m (the way in which the task was fulfilled by making use of the material; the structure, the make-up).[6]

These three aspects determine the overall division of grammar into three main parts, for which Noreen (1903: 50f; cf. also 1901: 3;) now gives, besides the native Swedish names, the artificially fabricated designations: *ljudlära*, or *fonologi*; *betydelselära*, or *semologi*; and *formlära*, or *morfologi*. As far as *semologi* is concerned, he (1904: 7) offers the following explanation:

> As a follow-up to the term *sememe*, I coined the term *semology* − cf. *phoneme*: *phonology* and *morpheme*: *morphology* − to replace the hitherto current, but clumsy "semasiology", the term used by the one who first insisted that the science of meaning should be introduced as a separate part of grammar, namely the German K. Reisig, and by many of his followers, including Pott.[7]

The science of meaning, according to Noreen's (1879: 28) estimation, was "the part of grammar that had been given the least attention".[8] Until the beginning of the 20th century, the time of the publication of the *Inledning* and of the first fascicles of *Vårt språk*, some important new developments would take place in that domain, including the foundation of "the new science of semantics" by Michel Bréal (1897), and the appearance − around 1883 − of thé corresponding French term *sémantique* (cf.: Gordon, 1982: 13ff; Nerlich, 1992: 6, 19). Noreen was *au courant*, and he found this

---

[6] Original: "Då språket väsentligen är − liksom kläder, boning och verktyg − en konstprodukt, så måste det för betraktelsen kunne erbjuda lika många och samma hufvudsynpunkter som hvarje annan sådan, nämligen m a t e r i a l e t s (det ämne, hvaraf konstprodukten förfärdigats), i n n e h å l l e t s (det ämne, som konstprodukten »föreställer» eller »behandlar»; den uppgift, som den har att »lösa», ändamålet) och f o r m e n s (det sätt, hvarpå uppgiften medelst det begagnade materialet lösts; strukturen, byggnadsstilen)."

[7] Original: "I anslutning till termen semem har jag sedan bildat termen semologi − jämför fonem: fonologi och morfem: morfologi − i stället för det hittils ofta brukade, men klumpiga »semasiologi», detta den term som användes af den, som först yrkade på införande af betydelselära såsom en självständig del af grammatiken, nämligen tysken K. Reisig, och af många hans efterföljare, hvaribland Pott."

[8] Original: "[…] den minst uppmärksammade delen af grammatiken."

new term better formed than Reisig's *semasiology*. Nonetheless, his own choice was different, and motivated by his concern for terminological consistency:

> [...] since the terms phonetics and phonology (science of sounds) must be unconditionally sharply distinguished, it is necessary [...] in consideration of the parallelism with phonology and morphology – the French themselves say "morphologie", not "morphique" or something like that – to replace semantics, though quite nice on the whole, with semology [...] (Noreen, 1904: 7).[9]

Apart from that, Noreen's dissatisfaction with both French semantics and German semasiology was conditioned by the fact that they only dealt with the history of meanings, curtailing thereby the range of the discipline and reducing it essentially to what he (1904: 8) called "etymological semology." In his own theory, this was only one part of the science of meaning, the other – and perhaps the more important one – being *descriptive* semology, concerned "with different kinds of Modern Swedish sememes, their mutual arrangement and their relationship to the linguistic forms, through which they are represented" (1904: 39).[10] This descriptive semology could be aptly divided, according to him, into two main parts, called in Swedish *kategorilära* (the analysis of categories) and *funktionslära* (the analysis of functions). The former must account for the semological system, or the system of semological categories in a certain language (i.e. psychological categories that have been given some linguistic expression), roughly equivalent to what some earlier authors had called "inner linguistic form." The task of the latter is to account for the interrelation between the semological categories and various linguistic forms whereby they are represented (cf. Noreen, 1901: 6; 1904: 25ff, 39).

It is not surprising that Noreen patterns his semology on the more developed phonology, whose central notion is that of *phoneme*. Curiously enough, Noreen's own contribution to the consolidation of phonology has somehow attracted much more attention from later linguists, than his attempt to improve the science of meaning, although these two branches of linguistics were treated in his doctrine on an equal footing. A number of distinguished scholars, including Björn Collinder (1938: 122-127), Bertil Malmberg (1964: 44), Aarni Penttilä (1971: 9), Giulio Ciro Lepschy (1972: 60), Tullio De Mauro (1984),[11] have drawn attention to the fact that both the term and the notion of phoneme are clearly present in Noreen's writings. It is true that Noreen (1903: 340), in speaking of qualitatively determined sounds as bearers of meaning differences, formulated a *notion* very similar to that of phoneme in modern

---

[9]Original: "[...] då termerna fonetik och fonologi (ljudlära) ovillkorligen måste strängt skiljas åt, så bör [...] på grund af parallelismen med fonologi och morfologi – äfven fransmännen säga »morphologie», icke »morphique» eller något dylikt – det annars rätt nätta semantik ersättas af semologi [...]."

[10]Original: "[...] för de olika slagen af nysvenska sememer, deras inbördes gruppering och deras förhållande till de språkformer, genom vilka de representeras [...]."

[11] Symptomatically, in a new dictionary of the Italian language edited by De Mauro (2000) the etymology of the word *semema* only reaches as far back as 1960, with a reference to the French *sémème*.

linguistics, but his own acceptation of the *term* "phoneme" departs very far from the present-day usage:[12]

> By o n e (in contrast to two or more) s p e e c h   s o u n d [...] I mean such a portion of the sound mass (a sound quantum) which can be produced by one and the same uniform "articulation", i.e. a fully fixed and regular arrangement of the vocal organs' movements that have been learnt and can be reproduced with mechanical readiness owing to practice. As for the   s e q u e n c e   o f s o u n d s , it denotes a larger or smaller group of sounds immediately following one another. After the example of the French,[13] I subsume both sound and sequence of sounds under the common term   p h o n e m e (sound mass, sound quantum). Accordingly, both single sounds, e.g. *s, k, r, i*, and sequences of sounds of smaller or larger size, e.g. *sk, skr, skri, skrik, skriker, skriker du, skriker du inte*, etc. are phonemes of different kind.[14]

If the use of the concept of phoneme implies a sharp distinction between phonology and its auxiliary science phonetics, to say nothing of the latter's auxiliary science acoustics, so the use of the concept of sememe implies an equally sharp distinction between semology and its auxiliary science philosophy of language, as well as psychology and logic (cf. Noreen, 1904: 50-51). The crucial question to be answered in this connection is the following one: "[...] what kind of mental content is it, that can function as the "meaning" of a linguistic expression of ours and thereby turn it into a sememe?" (Noreen, 1904: 13).[15] For the sake of brevity, Noreen (1904: 5) designates it with the term *idea*, characterized as "any sort of higher mental function of a theoretical kind," [16] and including, specifically:[17]

a)   P e r c e p t i o n , i.e. an immediate "intuition" [*this piece of chalk*];
b)   I n d i v i d u a l   r e p r e s e n t a t i o n ... i.e. a mediate intuition (a mental image, formed on the basis of past perceptions) [*our absent friend P. —* a mnemonic image; *your future husband —* an invented image];

---

[12] On the other hand, it is quite compatible with the meaning range of the Greek word φώνημα 'sound of the voice; speech; language'.
[13] In France, the term *phonème* was introduced in 1873 by A. Dufriche-Desgenettes, as the equivalent of the German *Sprachlaut*. It was subsequently taken up by Louis Havet and Ferdinand de Saussure.
[14]Original: "Med e t t (i motsats till två eller flere) s p r å k l j u d [...] menas så stor ljudmassa (så stort ljudkvantum) som frambringas medelst en och samma likformiga »artikulation», dvs. en fullt bestämd och regelbunden, inlärd och i följd af öfning med mekanisk färdighet företagen anordning af talorganernas verksamhet. Med l j u d f ö r b i n d e l s e åter menas en större eller mindre grupp af omedelbart på hvarandra följande ljud. Ljud och ljudförbindelse sammanfattar jag efter franskt föredöme under den gemensamma termen f o n e m (ljudmassa, ljudkvantum). Fonemer af olika slag äro sålunda såväl enstaka ljud, t. e. *s, k, r, i*, som ljudförbindelser af mindre eller större omfång, t. e. *sk, skr, skri, skrik, skriker, skriker du, skriker du inte* osv."
[15]Original: "[...] hvad är det för slags psykiskt innehåll, som kan uppträda såsom »betydelse» hos något vårt språkliga uttryck och sålunda göra det till ett semem?"
[16]Original: "[...] hvarje slags högre psykisk funktion af teoretisk art."
[17] In his classification of "ideas" (meanings), Noreen draws heavily on the achievements of 19th century psychology as they were reflected in his compatriot Frans Alexander von Schéele's (1853-1931) book *Det mänskliga själslifvet* (1896).

c)   Universal representation (general representation or "pattern"),
       i.e. a group of uniform representations, that change imperceptibly into
       one another ... [*a* (red, white, yellow etc.) *eglantine*]

d)   Concept, i.e. a conception (general and abstract, free of any image in
       its utmost manifestation) of what is similar and common for a group of
       representations, produced by the fixation of what is "essential," i.e.
       constant, for the universal representation, e. g. *three, triangle, the
       concept man, goodness.*

e)   Judgement (or group of judgements), i. e. the mental process
       through which an idea is perceived as linked with another idea, e.g. *han
       är adelsman* 'he is a gentleman', *gossen fryser* 'the boy is cold' (= has
       feelings of cold), *din dumbom!* 'you fool' (= you are a fool) (Noreen,
       1904: 16-17; cf. 1901: 4).[18]

Unlike the congenerous sciences, semology studies ideas only inasmuch as they
have "some linguistic garment" (Noreen, 1901: 3; cf. 1904: 5).[19] The term *sememe* is
used, precisely, to refer to all such ideas. Its definition is in fact very broad: "a definite
meaning content [...] in some linguistic garment, regardless of what linguistic form it
is clothed in" (Noreen, 1901: 3; cf. 1904: 6).[20] At this point already, it is important to
observe that although Noreen makes no explicit statement to this effect, his actual
terminological usage suggests rather clearly that sememes are treated in his doctrine as
double-faceted units, possessing an outer form of expression and a meaning.[21] Still, the
latter aspect evidently plays the leading part in semology and proves decisive in the
identification and classification of sememes, so that *triangle* and *trilateral rectilinear
figure* are considered to be one and precisely the same sememe, whereas the words
väsen 'noise' and *väsen* 'being' are different sememes (cf. Noreen, 1901: 3; 1904: 6f).
On the whole, Noreen (1904: 18) asserts, in line with very traditional argumentation,
going back to Plato[22] and Aristotle,[23] that the infinite variety of ideas exceeds by far

---

[18]Original:
a) Varseblifning, dvs. en omedelbar »åskådning» [*den här kritbiten*];
b) Individual-föreställning ... dvs. en medelbar åskådning (en inre bild, skapad på grundval af
forna varseblifningar) [*vår frånvarande vän P.* — en minnesbild; *din tillkommande* — en
fantasibild];
c)   Universal-föreställning (allmänföreställning eller »schema»), dvs. en grupp af likartade
föreställningar, som omärkligt öfvergå i hvarandra ... [*en* (röd, hvit, gul osv.) *törnros*]
d)   Begrepp, dvs. (en generell och abstrakt, i sin högsta potens rent bildlös) uppfattning af det
för en föreställningsgrupp lika och gemensamma, hvilket vinnes genom fixerande af det för
universalföreställningen »väsentliga», dvs. konstanta, t. e. *tre, triangel, begreppet människa,
godhet.*
e) Omdöme (eller grupp af dylika), dvs. den psykiska process, hvarigenom en idé uppfattas såsom
förbunden med en annan, t. e. *han är adelsman, gossen fryser* (= har köldförnimmelser), *din
dumbom!* (= du är en dumbom).
[19]Original: "[...] någon viss språklig dräkt [...]"
[20] Original: "[...] ett visst bestämmdt betydelseinnehåll [...] i språklig dräkt, oberoende av hvilken
språklig form det är iklädt".
[21] E.g., he speaks of an "expression" (*hvem*, i.e. 'whom') that constitutes, notwithstanding the variety
of its possible uses, "one and the same sememe" (cf. Noreen 1904: 38).
[22] Cf.: "[...] among the ancients [...] [there was no right] dividing genera into species [...] wherefore
there is no great abundance of names" (Plato, *Sophist*: 267d).

the number of linguistic forms of expression, and accordingly, each form, e.g. a morpheme, is typically polysemous, at least provisionally.

If meaning is so all-important for the identity of sememes, the problem of measuring differences in meaning comes to the fore in a very conspicuous way. To begin with, Noreen introduces the first big distinction between two fundamental types of meanings. According to him (1904: 19f.), a morpheme has:

> 1. A multitude of n o n c e (occasional) meanings, in conformity with the
>   change of circumstances each time it is used, e. g. *jag* 'I' = Mr A, Mr B,
>   Mr C, Mr D etc.
> 2. A single c o m m o n meaning,[24] obtained with the help of abstraction from
>   all the nonce meanings and therefore retaining only what is similar and
>   common to all of them, e.g. *jag* = the one who is speaking.[25]

We have to do with the same sememe, Noreen (1904: 22) affirms, as long as the differences in meaning concern solely the nonce meaning, not the common one. So, jag = Andersson and jag = Petersson are one and the same sememe. It would seem, then, that the sememes exist at the level of what is now described as "lexical", or "systemic" meanings, as opposed to the contextual ones. This conclusion proves, however, too hasty, as the following examples (cf.: Noreen, 1904: 22, 32) show:

(a)   *Hästen* är ett djur '*the horse* is an animal'.
        Hästen är ett fyrfotadjur '*the horse* is a quadruped animal'.
(b)   *Hästen* dog i natt '*the horse* died in the night'.

(c)   *En häst* har fyra fötter '*a horse* has four legs'
(d)   *En häst* går därborta på ängen '*a horse* is walking over there in the
        meadow'.

In Noreen's (1904: 22) opinion, *hästen* in (a) and (b) does not correspond to one sememe, but to two different sememes, "because they differ already by their concepts."[26] However, in this case, as well as in (c) and (d), neither of the meanings can be said to be "more common" than the other, and the difference between them has

---

[23] Cf.: "[...] since it is impossible to argue by introducing actual things under discussion, but we use names as symbols in the place of the things, we think that what happens in the case of the names happens also in the case of the things, just as people who are counting think in the case of their counters. But the cases are not really similar; for names and a quantity of terms are finite, whereas things are infinite in number; and so the same expression and the single name must necessarily signify a number of things" (Aristotle, *On Sophistical Refutations*: 165a).

[24] In the *Inledning*, Noreen (1901: 4) had preferred the term *general meaning (allmänelig betydelse)*, with the synonymous *common meaning (usuell betydelse)* given in brackets.

[25] Original:
  1.   En mångfald o c k a s i o n e l l a (tillfälliga) betydelser, alltefter omständigheterna olika för
        hvarje tillfälle, då det användes, t. e. *jag* = Hr A, Hr B, Hr C, Hr D osv.
  2. En enda u s u e l l betydelse, vunnen genom abstraktion ur alla de ockasionella betydelserna och
        därför upptagande blott det för dem alla lika och gemensamma, t. e. *jag* = den nu talande.

[26] Original: "[...] emedan dessa redan i fråga om sitt begrepp skilja sig åt [...]"

to be accounted for not by the "analysis of categories," but by the "analysis of functions." Noreen's (1904: 32) explanation runs as follows:

> Grammatical function or "linguistic use" refers [...] to a psychological category, represented in this case by a morpheme [...] Since most morphemes are polysemous, their function can very well vary in different cases, and so they have a certain, often very considerable functional latitude, e.g. *hästen*, the horse, that sometimes can signify some definite individual horse [...], and sometimes all horses, the whole equine race, the horse in general [...] [27]

The same functional approach underlies the division of sememes within the sentence structure:

> (e) *Hvem såg gossen? – hvem*: either Subject or Object-in-the-Accusative, i.e. 'who saw the boy?' vs. 'whom did the boy see?'
> (f) *Hvem gaf gossen brefvet? – hvem*: either Subject or Object-in-the-Dative, i.e. 'who gave the letter to the boy?' vs. 'to whom did the boy give the letter?'

Noreen's (1904: 38) comment distinguishes once more the rôles played by the analysis of categories and the analysis of functions: "[...] the expression *hvem* in the above examples [i.e., (e) and (f)] constitutes one and the same sememe in so far as it belongs to the same grammatical category, but different sememes in so far as it has different *functions* [...] "Meaning" in the large sense certainly comprises the function as well [...]."[28] Ultimately, the analysis of functions has the final say in linguistics,[29] although the analysis of categories proves very useful from the point of view of the method:

> If it is true that there is actually nothing but sentences in the real life of language, it does not prevent that, just as it can be appropriate in a manual of botany to consider the leaves, the roots, and other parts of a plant in isolation, in the same manner the scientific study of language may separate what in the life of language manifests itself only as united (Noreen, 1903: 46).[30]

---

[27] Original: "Med grammatisk funktion eller »språklig användning» menas nämligen den psykologiska kategori, som af ett morfem för tillfället representeras [...] Då de allra flesta morfemer ju äro polysema, så kunna de alltså ha olika funktion vid olika tillfällen, äga sålunda en viss, ofta mycket betydlig funktionslatitud, t.e. *hästen*, som kan betyda dels en viss bestämd individ af hästsläktet såsom i »hästen dog i natt», dels alla hästar, hela hästsläktet, hästen i allmänhet [...]."

[28] Original: "[...] uttrycket *hvem* i de anförda exempel är såtillvida ett och samma semem, att det tillhör samma grammatiska kategori, men såtillvida olika sememer, att det har olika f u n k t i o n [...] Till »betydelse» i vidsträckt mening hör naturligtvis också funktionen [...]"

[29] Noreen's concrete analyses are largely based on this approach, so that Benny Brodda (1973: 30) had every reason to consider him a forerunner of Charles Fillmore's theory of semantic cases.

[30] Original: "Om det [...] är sant, att det i det lefvande språklifvet egentligen icke gifves något annat än meningar, så hindrar detta dock icke, att, liksom det i en botanisk lärobok kan vara lämpligt att särskildt behandla blad, rötter o.a. växtdelar, äfven den vetenskapliga betraktelsen af språket äger rätt att söndra, hvad språklifvet endast har att uppvisa såsom förenadt."

Consequently, an important distinction arises, as far as semology is concerned, between two main types of sememes: bound ones and free ones. Noreen's special Swedish term for "bound sememe" is *glosa* (translated here as "locution"),[31] while for "free sememe", it is *yttrande*. The notion of locution is introduced in the following way:

> A l o c u t i o n is a subordinate or, if you prefer, bound sememe, i.e. a sememe whose meaning consists of a single element of thought (be it a perception, a representation, a concept or a collection of such), and which must appear in the real life of language in connection with others of the same kind, each single locution getting its specific meaning in and through the mental context it appears in. A special term for this concept has been lacking up to now, as the ambiguous expressions "ord" 'word' and "satsdel" 'clausal element', which have sometimes been used, can be neither recognized nor approved as such. [...] 'clausal element' or 'clausal constituent' is especially inappropriate because the locution can be an element (or constituent) not only of a sentence, but also of another locution; so e.g. the locutions *-s* in *faderns* or *o-* in *oviss* cannot be called clausal elements. Therefore, I propose that the term 'locution' should be adopted [...] (Noreen, 1904: 47; cf. 1901: 6).[32]

> From the formal point of view, locutions include:
> – combining elements: *-hona* or *kvinns-* 'female' e.g. in *räfhona* 'vixen', *kvinnsperson* 'woman',
> – suffixes: *-inna* (with the meaning 'feminine') in *herdinna* 'shepherdess'; *-s* (with the meaning 'belonging to') in *faderns* 'father's',
> – prefixes: *o-* (with the negative meaning) in *oviss* 'uncertain'; *van-* (with the meaning 'badly, wrongly') in *vanskapad* 'deformed'.[33]

As a purely semological concept, locution is identified as such only on the basis of its meaning. At the same time, "[...] locutions can differ very much as to their

---

[31] The Swedish word *glosa* generally signifies a word belonging to a foreign language, regarded, for instance, as something to be memorized.

[32] Original: "G l o s a är ett osjälfständigt eller, om man så vill, bundet semem, dvs. ett hvars betydelse utgöres af ett blott tankeelement (vare sig en varseblifvning, en föreställning, ett begrepp eller en samling af sådana), och som därför i det faktiska språklifvet måste uppträda i förbindelse med andra dylika, hvarvid hvarje särskild glosa erhåller sin specifika betydelse i och genom det tankesammanhang, i hvilket den förekommer. Detta begrepp har hittils saknat egen term, ty som sådana kunna de någon gång använda tvetydiga uttrycken »ord» och »satsdel» hvarken erkännas eller godkännas. [...] 'satsdel' eller 'satsled' är särskildt därför olämpligt, att glosan kan vara del (eller led) ej blott af en sats, utan äfven af en annan glosa; så t. e. kan man ej kalla glosorna *-s* i *faderns* eller *o-* i *oviss* för satsdelar. Jag föreslår därför fixerandet af termen 'glosa'[...]."

[33] It is interesting to note that Bloomfield (1950: 238) would express a very similar view in his *Language*: "[...] we find the sememe 'female of such-and-such male' expressed not only by the suffix *-ess*, but also by composition, as in *elephant-cow, she-elephant, nanny-goat*, and by suppletion, as in *ram: ewe, boar: sow*; some such pairs show inverse derivation, the male derived from the female, as *goose: gander, duck: drake*".

f o r m , i.e. consist of most disparate morphemes: word elements, words, collocations"
(Noreen, 1904: 48; cf. 1901: 6f).[34]

Noreen's (1904: 49) definition of utterance is fairly general:[35] "U t t e r a n c e   is
an independent or, if you prefer, a free sememe, i.e. a sememe whose meaning consists
of a whole train of thought of larger or smaller size; thus it is a closely connected, and
more or less self-sufficient whole."[36] Noreen (1904: 62; cf. 1901: 8) offers examples
of the attribute "larger or smaller size," which shed some light on what exactly he
means by "utterance." The examples he cites include *period, fragment* (or *section*),
chapter, part, volume, and *series.* So it becomes evident that he sets practically no
upper limit on the so-called compound utterance. In the opposite direction, a natural
limit is found, however, in the smallest possible instantiation of what can be called a
"train of thought" – a single judgement, expressed by a simple utterance, which can be
termed a *clause* (Sw. *sats*) from the morphological point of view, and a *sentence* (Sw.
mening) from the semological point of view (cf. Noreen, 1904: 41, 51, 57).[37]

In this connection, Noreen's psychologically tinged functional and semological
approach seems to become still more radical, as the conferral or non-conferral of
sentence status to a sememe is made depending on the mental reality of the judgement
that constitutes its meaning. His discussion of the possible ways of defining the
sentence is illuminating in this respect. In order to drive home his point, he refers to a
well-known definition by Hermann Paul (1846–1921):

> The sentence is the linguistic expression or symbol, denoting that the
> combination of several ideas or groups of ideas has been effected in the mind of
> the speaker; and is at the same time the means of reproducing the same
> combination of the same ideas in the mind of the hearer (Paul, 1891: 111).

Noreen's (1904: 56) objection to this definition is that the characterization of
the combination of several ideas as something that "has been effected" implies the
presence of an accomplished mental result, whose linguistic expression is not a clause,
but a locution. A judgement, on the contrary, is a mental process under way, the very
act of thinking, which can only be adequately described with the help of a form like "is
being effected."[38] Instead of Paul's definition, Noreen (1904: 53) prefers the one
proposed by Henry Sweet (1892: 155): "The sentence is a word or group of words
capable of expressing a complete thought or meaning."[39] Still, Paul's formulation does

---

[34] Original: "[...] glosorna till sin f o r m kunna vara af den mest olika beskaffenhet, dvs. utgöras af
de mest olikartade morfemer: orddelar, ord, ordfogningar."
[35] Noreen had not considered the notion of utterance in the *Inledning*, limiting his analysis there to
sentences and clauses.
[36] Original: "Y t t r a n d e är ett själfständigt eller, om man så vill, fritt semem, dvs. ett hvars betydelse
utgöres af en hel, större eller mindre tankegång, sålunda ett nära sammanhängande, i sig mer eller
mindre afslutadt helt."
[37] Noreen's approach correlates with the Aristotelian view of the λόγος ("phrase"), expressed in the
Poetics, ranging from a definition, which signifies "one thing," to the unity of the *Iliad,* as a
combination of several "phrases" (cf.: Aristotle, Poetics: 1457a).
[38] Cf.: Noreen, *Vårt språk* 5, p. 56.
[39] Elsewhere Sweet (1892: 19) specifies that it is "[...] a combination of a logical predicate with a
logical subject".

have a sort of utility, as it fits one type of sentence, namely, the so-called subordinate clause (in Swedish, *bisats*). Noreen's (1904: 57; cf. 1901: 8) treatment of subordinate clauses is based entirely on semological criteria while disregarding the formal ones:

> [...] a "subordinate clause" is semologically equivalent to a locution (cf. e.g. "Jag erkänner *att mannen är god*" 'I admit *that man is good*' and "Jag erkänner *mannens godhet*" 'I admit *man's goodness*'), wherefore I classify the subordinate clause as a type of "locution", rather than as a type of "sentence".[40]

The identity of meaning of the two cited examples leads Noreen (1904: 58) to the conclusion that the subordinate clause is "a former sentence, which assumed the function of a locution, but kept the appearance of a sentence, by being moulded in its most common form".[41]

If the outer form has any part to play in the framework of Noreen's semology, it is by fixing the lower limit of the resolution of trains of thought, when the morphologically simple elements are reached. On the other hand, the meanings of these elements are always "fairly compound,"[42] and should be accepted as such, "since language has no expressions whatsoever for simple sensations" (Noreen, 1904: 48; cf. 1901: 3).[43] Whatever more elementary units of meaning can eventually be arrived at by analytical dissection of mental contents, they bear no relevance for semology, because, lacking a linguistic form of expression, they cannot be regarded as facts of language. Just as the study of sound strings identifies the minimal functionally relevant "sound quantums" on the basis of their ability to produce differences in meaning, so the study of trains of thought identifies the minimal relevant "quantums of meaning" on the basis of their liability to acquire a definite phonetic expression. This seems to be, then, Noreen's version of the idea of isomorphism between the two separate but inseverable sides of language.

Although the honour of having created the *term* sememe should be given to Noreen, it cannot be overlooked that his *conception* of sememe (as "an ideal content expressed in some linguistic form") did not find its way into modern linguistics. One of the rare examples of some theoretical affinity (but hardly of direct borrowing) occurs in Goodenough (1956: 208), who defined the *lexeme* as "a morpheme, construction or phrase in those linguistic contexts where it has a single unpredictable significatum", the latter being identified, precisely, with a *sememe* (symptomatically, in this case we have to do with a scholar who is primarily an anthropologist, rather than a linguist).

One obvious reason for this lack of reception is the language *Vårt språk* was written in. However, such an explanation is not sufficient. The linguistic barrier did

---

[40] Original: "[...] en 'bisats' är semologiskt likställd med en glosa (jfr t. e. »Jag erkänner *att mannen är god*» med »jag erkänner *mannens godhet*»), och jag uppför därför bisatsen såsom ett specie af 'glosa', icke af 'mening'."
[41] Original: "[...] en f. d. mening, som öfvergått till att fungera som glosa, men behållit utseendet af en mening genom att vara stöpt i dennas vanligaste form."
[42] Original: "[...] ganska sammansatta [...]."
[43] Original: "[...] ty språket äger öfverhufvud inga uttryck för enkla förnimmelser (»sensationer» [...])."

not prevent, e.g., the recognition of Noreen's contribution to the development of phonology. Besides, one should not overlook the existence of a German compendium of Noreen's *magnum opus*, first published in 1923 and reprinted in 1975, which made a wider divulgation of his theoretical views, in principle, quite possible. A useful hint might be taken from the evaluation of Noreen's notion of sememe by Anton Joannes Bernardus Nicolaas Reichling (1898–1986), a Dutch scholar who got acquainted with Noreen's doctrine in this German version. In his most important book *Het woord* (1935), Reichling cites Noreen's (1923: 200; cf. 1904: 6, 23) identification of *Dreieck* and *dreiseitige geradlinige Figur* as the same sememe, and proceeds to make a pungent comment: "It is difficult to provide a clearer example of confusion between *thing[-meant]* and *meaning*" (Reichling 1967: 256, fn. 37).[44] Reichling himself is, on the contrary, very keen on differentiating the thing(-meant), which he (1967: 41) defines as "the "something" that we confront in our sensation or in our perception, the *not-I*",[45] from the meaning as such. The latter is a thing, too, but a thing *"sui generis"* (1967: 243), described as "a non-tangible *thing*, a thought, a concept",[46] or as "the not-I that we know *in the word*" (1967: 244).[47]

Apparently, Reichling altogether misses the point of discord between himself and Noreen, whose theory did account for the distinction in question, although in a different way (cf. his discussion of common vs. nonce meanings, or of the analysis of functions). In reality, Noreen's identification of *triangle* with *trilateral rectilinear figure* (quite unacceptable in Reichling's view) correlates with his idea that a sememe should indeed have "some linguistic garment", but without any specific requirements as to its concrete form. For Reichling (1967: 156), on the contrary, the crucial problem was that of the unity of the word:

> [...] it is not indifferent, whether we speak of "reference" and "referent" as applied to a "word" or to a "sentence". As an aspect of a "word", the "reference" is united with the "phonetic" side in an entirely special way, by the act of word-making, through which a p e r m a n e n t unit has arisen [...][48]

Unsurprisingly, in his ultimate explanation of what a word's meaning is, Reichling (1967: 358f) draws on one of the basic Saussurean notions:

> If we call to mind the word *boom* 'tree', with the question: "What does the word *boom* mean?," the actualized word, a s t h e p o i n t of d e p a r t u r e of o u r s c r u t i n y, can lack any further specification of its delimitation from everything that is not *boom*. But the word always possesses this minimum of

---

[44] Original: "Een duideliker voorbeld van verwarring tussen zaak en betekenis is moeilijk te leveren."

[45] Original: "Dat "iets", dat we in de waarneming of voorstelling tegenover ons stellen, het niet-ik [...]."

[46] Original: "[...] 'n onaanschouwelike zaak, 'n gedachte, 'n begrip [...]."

[47] Original: "[...] het niet-ik dat wij in het woord kennen."

[48] Original: "[...] het is niet hetzelfde bij een „woord" of bij een „zin" van „reference" en „referent" te spreken. De „reference" als moment van een „woord" is op een geheel eigenaardige wijze, door de act der woordmaking met de „phonetiese" momenten verenigd, waardoor een blijvende eenheid ontstaan is [...]."

definiteness. The word is, then, what *de Saussure* labelled with the term *valeur*. Its meaning consists in its thisness, in opposition to all the rest: it is a definite unit.[49]

Although Reichling has a number of emendations to make in Saussure's general theoretical framework (cf.: 1967: 22, fn. 5; 31; 307), his adoption of *valeur* as the basic criterion in the study of meaning, rooted as it is in the linguistics of *langue*, seems to explain his cecity to Noreen's argumentation. The latter, indeed, regarded the study of the "living" language, in the first place, as what would come to be called the linguistics of *parole*. Accordingly, his sememes are not to be differentiated from other sememes within lexical paradigms, as in Reichling and many others, but rather within syntagmatic strings – which had been a typical perspective of linguistic analysis since antiquity (cf.: Rehn, 1986: 64). With the advent of the Saussurean turn, Noreen's notion of sememe had no chances of survival. The fact remains that it is at home in a domain where the "classical" structural semantics of Saussurean inspiration encounters most serious difficulties: that of discourse analysis.

References

Aristotle. 1927. Poetics (translated by William Hamilton Fyfe). In: Aristotle. *The Poetics. "Longinus" on the Sublime. Demetrius on Style.* London: Heinemann; Cambridge (Mass.): Harvard University Press. 1447–1462.

Aristotle. 1955. On Sophistical Refutations (translated by Edward Seymour Forster). In: Aristotle. *On Sophistical Refutations. On Coming-to-Be and Passing-Away. On the Cosmos.* London: Heinemann; Cambridge (Mass.): Harvard University Press. 165–184.

Bidu-Vrănceanu, Angela and Narcisa Forăscu. 1984. *Modele de structurare semantică. Cu aplicaţii la limba română.* Timişoara: Facla.

Bloomfield, Leonard. 1950. *Language.* London: George Allen & Unwin Ltd.

Bréal, Michel. 1897. Une science nouvelle. La sémantique. *Revue des deux mondes.* 67 (141): 4, 807–836.

Brodda, Benny 1973. *(K-)overta kasus i svenskan.* Stockholm: Institute of Linguistics.

Collinder, Björn. 1938. Lautlehre und Phonologismus. In: *Actes du quatrième Congrès international des linguistes, tenu à Copenhague, du 27 août au 1er septembre 1936.* Ed. by Kaj Barr, Viggo Brøndal, Louis Leonor Hammerich & Louis Hjelmslev. Copenhagen: Munksgaard. 122–127.

Coseriu, Eugenio and Horst Geckeler. 1974. Linguistics and Semantics. Linguistic, Especially Functional, Semantics. In: *Current Trends in Linguistics* 12,

---

[49] Original: "Roepen we ons 't woord *boom* voor de geest, met de vraag: "Wat betekent 't woord boom?," dan kan aan dat geactueerde woord, als uitgangspunt van onze beschouwing, elke andere praecisering dan zijn afgrensing tegen alles wat niet *boom* is, ontbreken. Maar, dit minimum van bepaaldheid heeft 't woord altijd. Het woord is dan, wat *de Saussure* met de term *valeur* bestempeld heeft. Zijn betekenis bestaat in z'n dit-zijn, in tegenstelling met al het andere: het is 'n duidelike eenheid."

Linguistics and Adjacent Arts and Sciences. Ed. by Thomas Albert Sebeok. The Hague & Paris: Mouton. 103–171.

Crystal, David. 1987. *The Cambridge Encyclopedia of Language.* Cambridge, etc.: Cambridge University Press.

DeMauro, Tullio. 1984. Appendice. Noreen et Saussure. In: Ferdinand de Saussure. *Cours de linguistique générale.* (édition critique preparée par Tullio de Mauro). Paris: Payot [Italian edition first published in 1967]. 390–394.

----- (ed.). 2000. *De Mauro. Il dizionario della lingua italiana.* Torino: Paravia.

Filipec, Josef and František Čermak. 1985. *Česká lexikologie.* Prague: Academia.

Goodenough, Ward Hunt. 1956. Componential Analysis and the Study of Meaning. *Language.* 32 (1): 195–216.

Gordon, William Terrence. 1982. *A History of Semantics.* Amsterdam & Philadelphia: Benjamins.

Greimas, Algirdas-Julien. 1966. *Sémantique structurale. Recherche de méthode.* Paris: Larousse.

Heger, Klaus. 1964. Die methodologischen Voraussetzungen von Onomasiologie und begrifflicher Gliederung. *Zeitschrift für romanische Philologie.* 80: 486–516.

Huddleston, Rodney. 1974. Componential analysis: the sememe and the concept of distinctiveness. *The Canadian Journal of Linguistics.* 19 (1): 1–17.

Lavorel, Pierre Marie 1975. *Éléments pour un calcul du sens.* Paris: Dunot.

Lepschy, Giulio Ciro 1972. *A Survey of Structural Linguistics.* London: Faber and Faber.

Lotz, John. 1954. Plan and Publication of Noreen's *Vårt Språk. Studia Linguistica.* VIII: 82–91.

Lounsbury, Floyd Glenn. 1956. A Semantic Analysis of Pawnee Kinship Usage. *Language.* 32 (1): 158–194.

Löwkrantz, Gottfrid. 1901. Sommarkurserna i Uppsala. Inledning till modersmålets betydelselära. *Svensk läraretidning.* 20 (36): 603–604.

Lyons, John. 1977. *Semantics* 1. Cambridge, etc.: Cambridge University Press.

Malmberg, Bertil. 1964. *New Trends in Linguistics: an orientation* (translated from the Swedish original by Edward Carney). Stockholm: Naturmetodens språkinstitut.

Meier, Georg Friedrich. 1964. Semantische Analyse und Noematik. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung.* 4: 581–595.

Nerlich, Brigitte. 1992. *Semantic Theories in Europe 1830–1930: From Etymology to Contextuality.* Amsterdam & Philadelphia: Benjamins.

Noreen, Adolf. 1879. Något om ord och ordklasser. *Nordisk tidskrift för vetenskap, konst och industri.* 2: 18–29.

-----. 1901. *Inledning till modersmålets betydelselära.* Uppsala: Almqvist & Wiksell.

-----. 1903. *Vårt språk: nysvensk grammatik i utförlig framställning* 1: *Allmänn inledning.* Lund: Gleerup.

-----. 1904. *Vårt språk: nysvensk grammatik i utförlig framställning* 5: *Betydelselära (Semologi).* Lund: Gleerup.

-----. 1923. *Einführung in die wissenschaftliche Betrachtung der Sprache: Beiträge zur Methode und Terminologie der Grammatik* (Vom Verfasser genehmigte und durchgesehene Übersetzung ausgewählter Teile seines schwedischen Werks *Vårt språk* von Hans Wolfgang Pollak). Halle: Niemeyer.

Oleksenko, Olena. 1999. *Typolohija hramatyčno zvjazanyx značen' sliv*, Kharkiv, s.n.

Paul, Hermann. 1891. *Principles of the History of Language* (translated by Herbert Augustus Strong). London: Longmans, Green, and Co.

Penttilä, Aarni. 1971. Die "Emische" Sprachwissenschaft. *Studia Linguistica.* XXV: 9–20.

Plato. 1875. Sophist. In: *The Dialogues of Plato* IV (translated by Benjamin Jowett). Oxford: Clarendon. 216a–268d.

Pottier, Bernard. 1963. Du très général au trop particulier en analyse linguistique. *Travaux de Linguistique et de Littérature.* I: 9–16.

-----. 1964. Vers une sémantique moderne. *Travaux de linguistique et de littérature.* II (1): 107–137.

----- (ed.). 1973. *Le langage.* Paris: Centre d'Étude et de Promotion de la Lecture.

Pottier, Bernard, Albert Audubert, and Cidmar Teodoro Pais. 1973. *Estruturas Lingüísticas do Português.* São Paulo: Difusão Europeia do Livro.

Rehn, Rudolf. 1986. Zur Theorie des Onoma in der griechischen Philosophie. In: *Sprachphilosophie in Antike und Mittelalter. Bochumer Kolloquium, 2.-4. Juni 1982.* Ed. by Burckhardt Mojsisch. Amsterdam: Grüner. 63–119.

Reichling, Anton. 1967. *Het woord. Een studie omtrent de grondslag van taal en taalgebruik.* Zwolle: W. E. J. Tjeenk Willink.

Schéele, Frans Alexander von. 1896. *Det mänskliga själslifvet*, Stockholm, s. n.

Sweet, Henry. 1892. *English Grammar: Logical and Historical* 1, Oxford: Clarendon Press.

Contact details: serhijwakulenko@hotmail.com

Claimed and Unclaimed Sources of *Corpus Linguistics*

Jacqueline Léon
CNRS, Université Paris 7, France

The promotion of the term *Corpus Linguistics* in the 1990s has marked an important milestone in the attempt of making corpus works a new mainstream discipline within Language Sciences. Since an international conference held in 1991 gathering British, Dutch, Swedish and Norwegian linguists (proceedings in Svartvik (ed) 1992), the researchers of the domain have strengthened their position by the publication of many collective books and text books and the creation in 1996 of an international journal *The International Journal of Corpus Linguistics*. So in 2002, Geoffrey Leech (b. 1936), a leading figure in corpus research, could speak of 'the corpus linguistic world as a well-established research community' (Leech, 2002:167). At the same time as they were giving themselves a name, dating the first occurrence of the term back to 1984 (Aarts and Meijs, 1984), these linguists attempted to provide the new domain with historical legitimacy. It is this issue that I would like to address in my paper in order to see under what conditions, but also at what cost, a common history has been claimed in order to found a new linguistic stream. In particular, it will be shown how the actors have retrospectively built their own history by overstating or forgetting some events, facts or methods.

First it is to be noted that what is called 'Corpus Linguistics' covers various heterogeneous fields ranging from lexicography, descriptive linguistics, applied linguistics – language teaching or Natural Language Processing – to domains where corpora are needed because introspection cannot be used, such as studies of language variation, dialect, register and style, or diachronic studies. The sole common point to these diverse fields is the use of large corpora of texts or spontaneous speech, available in machine-readable form – often including statistical or probability methods but not systematically. Corpus investigations involve inductive method instead of hypothetico-deductive method, meaning that data-driven analyses are preferred to rule-driven ones. The point here is why one should gather such a diversity of approaches under a single term. Besides, why define a specific domain when, since linguistics indisputably remains an empirically based scientific area, any linguist is a potential user of corpora.[1]

In his paper entitled "Corpora and theories of linguistic performance," Leech (1992) promotes computer-based corpus research as a new paradigm of linguistics, denying that it may be regarded as a mere technique or method: 'I wish to argue that computer corpus linguistics defines not just a newly emerging methodology for studying language, but a new research enterprise, and in fact a new philosophical approach to the subject.' (Leech, 1992: 106-107) and lists its main features:

　　- Focus on linguistic performance, rather than competence;

---

[1] See for example Fillmore's paper (1992) in one of the corpus linguistics text books, which is an attempt to reconcile intuitive methods (armchair linguistics) with corpus linguistics. See also the French Journal *Corpus* which presents studies using both corpus and hypothetico-deductive methods.

- Focus on linguistic description rather than linguistic universals;
- Focus on quantitative, as well as qualitative models of language;
- Focus on a more empiricist, rather than a rationalist view of scientific inquiry.

Leech insists that each of these features highlights a contrast between the Corpus Linguistics paradigm and the Chomskyan paradigm (and dedicates a long development to it in his paper). In other words, this set of propositions is presented as an anti-Chomskian paradigm.

Let us now examine the common history which is now so largely spread among the researchers and text books that it can be said that a real historiography has been retrospectively set up in the 1990s.

1. The common story

The first versions of the story appeared in Leech's contribution to the studies published in honour of Jan Svartvik (1991), and in two sets of works regarded as significative of the resurgence of the domain claimed at the beginning of the 1990s: Svartik's collective book of 1992 and the special issue of the journal *Computational Linguistics* about using large corpora published in 1993, introduced by a historical overview (Church and Mercer, 1993). Note that, in the area of Natural Language Processing, the domain, called either *Computational Linguistics Using Large Corpora* or *Corpus-based Natural Language Processing*, had not been dubbed by any specific term at that time.[2]

Leech (1992) gives a simplified version of the story. In the 1940s-50s, corpora were flourishing among American structuralists, for whom 'a corpus of authentically occurring discourse was the thing that the linguist was meant to be studying' (1992: 105). Afterwards corpus linguistics went to sleep for twenty years and only came back in the 1980s with the increasing power of computers and the availability of very large corpora. Chomsky's criticisms of the 1950–60s are put forward to account for the decline of corpus linguistics: 'the impact of Chomskyan linguistics was to place the methods associated with CCL [Computer Corpus Linguistics] in a backwater, where they were neglected for a quarter of a century' (Leech, 1992 :110).[3]

This version of the story takes a more general move with the resurgence of empiricism against rationalism in the 1990s claimed both by Leech (1992) and Church and Mercer (1993): Computer Corpus Linguistics, as well as Corpus-based Natural Language Processing, are claimed to be a rediscovery of empirical and statistical methods popular in the 1950s, in particular the application of Shannon's information theory. After the 1950s empiricism declined, while rationalism became dominant in the areas of linguistics and artificial intelligence, marked by Chomsky's criticism of n-grams in *Syntactic Structures* (1957) and by Minsky and Papert's criticism of

---

[2] At present, it seems that in Natural Language Processing too the term Corpus Linguistics has been adopted. However other terms have appeared such as "Statistical Natural Language Processing" (Manning and Schütze, 1999) or "Probabilistic Linguistics" (Bod, Hay and Jannedy, 2003).
[3] See the same argument in Leech (1991): 'The discontinuity can be located fairly precisely in the late 1950s. Chomsky had effectively put to flight the corpus linguistics of the earlier generation.' (1991: 8).

Perceptron neuronal networks in 1969.[4] In addition to the increase in computer power and data availability, Church and Mercer give various reasons for the resurgence of interest in these methods in the 1990s. It was first in the area of Speech Recognition that stochastic methods, based on Shannon's model, reappeared in the 1970s when knowledge-based and rule-based methods became unsatisfactory and were given up. This change of method in speech recognition led computational linguistics to adopt probabilistic methods, notably preference-based parsing and lexical preference parsing.[5] Both versions of the story share the fact that they distinguish two corpus periods, the 1950s and the 1990s, and that, in between, corpus linguistics is said to have vanished essentially because of Chomsky.

A slightly different version had been proposed by Leech in 1991, pointing out the apparition of a second intermediary generation of corpus at the beginning of the 1960s: Randolph Quirk's Survey of English Usage (SEU) and Kucera and Francis's Brown corpus, presented as 'the founders of a new school of corpus linguistics, little noticed by the mainstream' (Leech, 1991: 8). In this version, however, it is not mentioned that the SEU predated and influenced the Brown corpus. On the contrary, the Brown corpus is considered the first real computerized corpus. What is more, this pioneer status of the Brown corpus has been taken up by many text books and collections of articles and is now widely shared among the actors of the field.[6] Let us now examine the key features of the story:
- the anteriority of the Brown corpus;
- the discontinuity of corpus design over 30 years;
- Chomsky's arguments against corpora and statistics.

2. The anteriority of the Brown corpus

The pioneer status of the Brown corpus rests on several assumptions: it had no precursor; it was the first computerized corpus and the first freely available corpus; it was supposed to favour general linguistic investigations and not just frequency counts of vocabulary.

The Brown corpus's authors claimed that it had no precursor. The front flap of Kucera and Francis's 1967 book asserts: 'The standard corpus of present-day edited American English prepared at Brown is the first and so far the only such collection of data in English that has been carefully selected by random sampling procedures from a well-defined population and that is completely synchronic, containing samples published in the US during a single calendar year (1961).' No other corpus is ever mentioned in the book, and since its publication it has been referred to as "the Brown corpus". This view was resumed thirty years later: 'The beginning of it all was the making of the Brown corpus, "a standard sample of present-day English for use with

---

[4] See Rosenblatt, 1958.
[5] See also the recent investigations of language probabilistic properties in Bod and al. (2003).
[6] Several text books have adopted this version since the 1990s. See for example :
Oostdijk and Haan (eds.), 1994; McEnery and Wilson, 1996; Garside, Leech, MacEnery, 1997; Simpson and Swales (eds.), 2001.

digital computers." In the next two decades it was to be followed by a string of successors...' (Svartvik, 1992: 7).

The only acknowledged sources are quantitative analyses of language and literary genre studies, especially word-frequency distribution in various languages. Thus Kucera and Francis's references essentially concern statistical works, namely Yule's and Herdan's works. In this respect, their work was not very original since word-frequency counts were flourishing at the time.[7] It should be added that Kucera was a Slavist of Czech origin, acquainted with the Prague School and its tradition of genre studies; for that matter, one of their references mentions a talk on statistics and genres given in 1966 at Brown University by Lubomir Dolezel (b.1922), one of Kucera's Czech compatriots. Not surprising, then, that the sampling of the corpus rested on genre categories and that the first studies concerned statistical genre studies.

Yet it can be shown that the Brown corpus had other sources: it resulted from a joint idea of one of Firth's pupils, Randolph Quirk, and the American Germanist Freeman Twaddell.[8] Randolph Quirk (b. 1920) decided in 1959 to devise a corpus of both spoken and written British English, the Survey of English Usage (SEU) at University College London (Quirk, 1960). The corpus was planned to be machine-readable, and though it was only computerized in 1989, Quirk took a programming course in order to achieve this purpose in the sixties.[9] The corpus was collected with the ultimate aim of supplying material for the writing of a grammar, the first version of which was published in 1972, *A Grammar of Contemporary English*, co-written by Quirk, Greenbaum, Leech and Svartik.[10]

In various respects the SEU belonged to the British tradition. Besides the fact that Quirk presented his scheme to the Philological Society and that he published his first paper on the SEU in its *Transactions,* the first investigations were for the most part on prosody and phonetics, that is following the phonetic tradition of Daniel Jones and J.R. Firth.

Freeman Twaddell (1906–1982), a specialist of German literature and phonology, created two departments at Brown University in Providence (Rhode Island), one of linguistics and one of Slavic studies in the early 1960s. In 1955, he invited Henry Kucera (b.1925) and in 1962 Nelson Francis (1910–2002) to join these departments and to participate in his corpus project.[11] The corpus was assembled at

---

[7] A glimpse of the extent of the area can be caught from Pierre Guiraud's *Bibliographie critique de la statistique linguistique*, published in 1954, where more than 240 references are exclusively dedicated to word counts.
[8] It should be noted that Firth's filiation is never mentioned by the Brown corpus researchers, whereas Firth is considered one of the precursors of corpus linguistics by M.A.K. Halliday, John Sinclair and their followers, and several recent corpus studies mention the notion of collocation as one of the touchstones of corpus linguistics (Léon, in preparation).
[9] The data were recordings amounting to 30,000 words (over three hours) of spontaneous English speech in the form of discussions between a total of 31 educated British adults (Crystal and Quirk, 1964). By 1991, over 200 publications had used material from the Survey Corpus, either in its original slip form or in its later computerized form (Altenberg, 1991).
[10] However, this grammar cannot be said strictly speaking to be "data-driven". As Sinclair (1991) points out, occasional reference is given to the SEU, and only a few examples had been extracted from the corpus.
[11] Kucera was a specialist of Czech phonology and became a computational linguist in order to achieve the comparison of the phonological similarities of three languages: Russian, Czech and German

Brown University during 1963-64; its analysis was performed during 1965-66 and the results published in 1967.

The connections between the SEU and the Brown corpus were strong and the anteriority of the SEU cannot be denied. During a scholarship in the United States in 1951, Quirk met Twaddell at Brown University and followed the teaching of Charles Fries at Ann Arbor. The planning of the SEU therefore owes much to Fries's empirical method of working on the syntax of spoken language (Quirk, 2002). In 1962, Nelson Francis obtained a scholarship from the Ford foundation to work on the SEU at UCL. Finally, in 1963, Randolph Quirk was present at the original conference that agreed on the specifications of the Brown corpus and where the major decisions were made (Kucera and Francis, 1967: xx).

Let us move on with the claim concerning the linguistic aims of the Brown corpus. Its ambition was to be a 'reservoir of linguistic usage in a form (computer tape) that makes it relatively easy to extract exhaustively all available specimens of a given word or describable grammatical item' (Twaddell's foreword in Kucera and Francis, 1967: v).

In the event, when we study Kucera and Francis's book, it is entirely focused on statistical studies. As they point out in their introduction: 'The main objective of this book is the presentation of lexical and statistical data about the Corpus' (Kucera and Francis, 1967: xvii); 'The bulk of the book [pp. 5-274] comprises two frequency lists of the words in the Corpus, the first in descending order of frequency and the second alphabetical. The rest [pp. 275-430] comprises a number of tables and graphs resulting from various counts, calculations, and studies and two essays analyzing some of the results' (Kucera and Francis, 1967: xxi-xxii). The book was in fact devoted to the distribution of frequent words, as well as word length and sentence length distribution, carefully carried out on sampled texts and genres. It was only later that the Brown corpus was used for grammatical investigations, once Quirk and his colleagues had already published several studies on grammar and prosody based on the exploration of the SEU (Quirk and Mulholland, 1964; Quirk and al., 1964; Quirk and Crystal, 1966).

Thus, at the end of the 1960s, the Brown corpus was used for exactly the same purpose as were quantitative data at that time, that is, statistical studies of vocabulary. Looking at Altenberg's bibliography (1991) listing the publications using English computer corpora, it can be seen that by 1970 more papers (10) using the SEU for grammatical or prosodic investigations had been published than using the Brown corpus for grammatical investigations (4 papers).[12]

The Brown corpus's main quality rested on its method of sampling. Its even sample size of 2,000 words helped to make the corpus statistically sound and was an

---

(Kucera, 1963; Kucera and Monroe, 1968). In particular, he had to write programs to deal with his phonological frequency data. He taught computational linguistics to his colleague Nelson Francis who had been trained in philology and dialectology (Francis, 1998; Kucera, 1998).

[12] Note that Altenberg lists the SEU among computer corpora even though it was not computerized at that time.

advantage for comparison.[13] However, as Sinclair (1991) pointed out, the divisions of the corpus into genres, settled on intuitive criteria, were less reliable. Moreover, following Gellerstam (1992) commenting on the Brown corpus, it can be said that the sampling method was more suited to producing quantitative results than conducting general linguistic investigations.[14]

The next claim concerns the anteriority of the Brown corpus as the first computerized corpus. This again has to be mitigated. There was at least one computerized predecessor; the computerization of the Trésor de la Langue Française had begun before the Brown corpus was planned. In 1957 a conference about 'Lexicologie et lexicographie françaises et romanes' took place in Strasbourg to study the faisability of a Dictionary of modern and contemporary French (1789–1960) based on a computerized Thesaurus or "Trésor de la Langue Française" (later called TLF), that is a corpus of 1350 literary or technical books written from 1789 to 1960. Subsequently, the Center of TLF was created by the CNRS (Centre National de la Recherche Scientifique) in 1960 (CNRS, 1961).[15]

Thus the anteriority of the Brown corpus cannot be settled on the computational level. However it is true, that unlike the TLF, the Brown corpus was made available immediately, and a tape provided to any researcher who asked for it.[16]

There is another predecessor which has been completely forgotten by corpus historiography, namely the Rand corpus. This corpus was created by the Rand Corporation group of Machine Translation in Los Angeles, led by David Hays (1928-1995) from 1960 to 1969. They developed empirical methods in Machine Translation as early as 1949. The method was to derive a dictionary and a grammar from the corpus and to test them on a new sample of the corpus to expand them. This method of building data-driven grammars and dictionaries was much more ambitious than the frequency explorations of the Brown corpus. Thus as early as 1959, the Rand corpus, comprising more than 200 Russian articles in Physics and Mathematics and more than two hundred thousand running words, was made available in the form of punched cards for the use of researchers.[17]

To return to the Brown corpus, we can see that it was not the first available computerized corpus, and that data-driven grammars date back to the Rand corpus.

---

[13] The corpus, of more than one million words, comprised 500 samples of 2000 words each. Fifteen categories (genres) were represented, from sports, scientific journals, and popular fiction to philosophical discussion.

[14] 'By looking at sampling principles, you can see that the focus was on obtaining quantitative data (frequencies of words, constructions, morphemes, graphemes) rather than on compiling a range of corpora useful for different purposes. An important point was to make the corpus so diversified that no individual text could possibly distort the frequency figures.' (Gellerstam, 1992: 152).

[15] In 1970, when the editorial work on the dictionary actually began, the computerized corpus comprised more than 80 millions of running words. It was completed by indexes, frequency counts and concordances (Martin, 2000).

[16] Leech (1991) mentions the TLF as a pioneer corpus of written texts, but not in connection with the second wave of corpora which he says appeared in the 1970s.

[17] It should be remembered that empirical methods had been sternly criticized by Machine Translation researchers themselves, first of all by Yehoshua Bar-Hillel (1915–1975) in his report of 1960 which initiated the decline of Machine Translation (Bar-Hillel, 1960).

However, it can be understood that the Machine Translation filiation, made infamous by the ALPAC report of 1966, has been forgotten.[18]

3. The discontinuity of corpora design over 30 years

Now let us examine another point: the claimed discontinuity of corpus linguistics during 20-30 years, between the corpus work of the American structuralists in the 1950s and the revival of corpora in the 1990s. Actually the accounts diverge on this point. Leech (1992) asserts there was a break of twenty years, while in his version of 1991 he mentions the appearance of a second generation of corpora in the early 1960s. Svartvik (1992) points out that the Brown corpus of the late 1960s was followed by a string of successors in the 1970s.

In fact the publication of many text books on computerized corpora during the 1970s-80s attests to this continuity. In particular, Aarts and Meijs (1984) comprises papers about the successors of both the Survey of English Usage and the Brown corpus which had been developed in the 1970s in collaboration with Swedish, Norwegian and Dutch universities. Let us only mention the first most famous ones: in 1975, the London-Lund Corpus of Spoken English (LLC); in 1978, the Lancaster-Oslo-Bergen Corpus of British English (LOB); in 1987, the Collins Birmingham University International Language Database (COBUILD) which is also a metonymic name for the Collins COBUILD English Language Dictionary published in 1987.

To conclude this point, it can be said that there was no break between the 1960s and the 1980s or 1990s in the production of corpora. Incidentally, the text books published before the 1990s never mention the priority of the Brown corpus, nor any attempt made by Chomsky to stop corpus development.

4. Chomsky's arguments against corpora and statistics

This remark leads to our last point. According to current corpus linguists, Chomsky stopped corpus linguistics in the 1950s, so that the pioneering Brown corpus appeared in a very hostile context:

> The impact of Chomskyan linguistics was to place the methods associated with CCL [Computer Corpus Linguistics] in a backwater, where they were neglected for a quarter of a century (Leech, 1992: 110).

> The discontinuity can be located fairly precisely in the late 1950s. Chomsky had effectively put to flight the corpus linguistics of the earlier generation. (Leech, 1991: 8).

---

[18] Since the ALPAC report which put an end to MT research in the USA and in the rest of world, Machine Translation has been included in Computational Linguistics, and the first MT experiments, innovating though they might be, have hardly been mentioned (see Léon, 1999; Cori & Léon, 2002).

> The Brown Corpus was significant not only because it was the first computer
> corpus compiled for linguistic research, but also because it was compiled in the
> face of massive indifference if not outright hostility from those who espoused
> the conventional wisdom of the new and increasingly dominant paradigm in US
> linguistics led by Noam Chomsky. (Kennedy, 1998: 19).

These statements involve two assumptions. First, that the early computerized
corpora, and particularly the Brown corpus, should be regarded as the revival or the
continuation of American Structuralists' conception of corpora. Second, that
Chomsky's criticisms equally concerned every statistical model and were virulent
enough to stop any form of corpus research.

Let us examine the first assumption. In post-Bloomfieldian linguistics, a theory
should aim at a systematic taxonomy of linguistic elements (distributional classes)
from a corpus of observed data through discovery procedures. The critics of
structuralism, most notably Chomsky, argued that these procedures yielded no more
than a static inventory of signs, devoid of any significance and not allowing any
theoretical explanation. The description obtained by this method was limited to the
data which had been collected and led to no insight into the nature of language.[19]

Now it can be seen that the idea of corpus at work in the Brown corpus does not
match the American structuralist approach. The authors and users of the Brown corpus
conceived of the corpus as a set of observed utterances from which frequency counts
could be investigated. No idea here of taxonomy, nor of discovery procedures.
Whereas statistics were used by American structuralists to predict which sentences
belonged to the corpus and which ones did not, the Brown corpus users aimed to
compare frequency counts between genres, or to test general statistical models on
vocabulary.

In fact, the links of the Brown corpus with American structuralism were rather
loose. As Falk (2003) has shown, none of Twaddell's contemporaries included him
among the structuralist linguists.[20] What is more, Twaddell does not seem to have
conceived of the Brown corpus as a taxonomic machine. In his foreword to Kucera
and Francis's book, he points out the advantages and disadvantages of the use of a
corpus in linguistics without claiming strong theoretical views:

> The advantages and disadvantages of basing linguistic statements on a specific
> corpus are familiar. A corpus protects against gross lapses of recall to which
> introspection is liable. Statistical statements depending on subjective judgments
> are unreliable in the extreme. On the other hand, a corpus of manageable size
> will underrepresent some units and structures that introspection can supply and
> specify adequately. (Twaddell's foreword in Kucera and Francis, 1967: v)

---

[19] Recall that Hockett's views on corpora, as early as 1948, involved the notions of infiniteness of
language and of projection (Hockett, 1948).
[20] In his book dedicated to the evaluation of post-Bloomfieldian achievements and Chomsky's work,
Hockett ranks Twaddell among those 'whose training had been, for better or worse, somewhat freer of
Bloomfield's influence' (1968: 18).

As to Quirk, he devoted more discussion to theoretical issues and to the divergences between Chomsky and the post-Bloomfieldians. In particular, Quirk and Svartvik (1966) took part in the debate, widespread among the linguists and psychologists of those days, on Chomsky's approach of competence versus performance, and grammaticality versus acceptability. Though Quirk and Svartvik supported performance and acceptability against competence and grammaticalness, they proposed an experiment for establishing degrees of acceptability in English sentences, which they used later in their grammar (Quirk and al., 1972).

Now, let us briefly examine the early arguments put forward by Chomsky against corpora in "Three Models of grammar" (1956), taken up in *Syntactic Structures* (1957), in the discussion between Chomsky and American structuralists which took place at the University of Texas (1958 [1962]), and finally in Miller and Chomsky (1963).

Note however that ten years separated the publication of *Syntactic Structure* and the Brown corpus and that Chomsky's early criticisms of the use of corpora in syntax could not concern this corpus, nor any computerized corpus since they did not exist yet at that time. Actually, Chomsky did not attack word frequency counts strictly speaking, since his main criticisms concern the use of probabilities, essentially Markov's model and Shannon's information theory, in syntactic analysis. As information theory and discovery procedures were widely debated in the 1950-60s, Chomsky's criticisms were not isolated and their reach and strength should not be overstated.

Chomsky argues that any particular corpus of utterances obtained by linguists in their fieldwork cannot be identified as the set of grammatical sentences, inasmuch as the notion of grammaticality involves those of projection, infiniteness and ideal speaker:

> Any grammar of a language will *project* the finite and somewhat accidental corpus of observed utterances to a set (presumably infinite) of grammatical utterances. In this respect, a grammar mirrors the behavior of the speaker who, on the basis of a finite and accidental experience with language, can produce or understand an indefinite number of new sentences. (Chomsky, 1957: 15)

When invited by Archibald A. Hill (1902–1992) in 1958 to present his linguistic model at the University of Texas, Chomsky (1962) addresses this issue a little differently, arguing that any natural corpus is skewed. If generated, it will produce non-sentences or conversely be incomplete and not provide every grammatical sentence.[21] In addition the description would be reduced to a mere list without any explanatory hypothesis.

Moreover, Chomsky claims that grammaticality cannot be identified with high statistical approximation, and criticizes descriptivists' suggestion of replacing possible sentences by highly probable sentences and impossible sentences by low probability

---

[21] This argument has been reported by Leech (1991, 1992), as well as Chomsky's issues developed since 1965, performance / competence and I-language / E-language, which we do not address in this paper.

sentences (Chomsky, 1957). As a matter of fact, grammaticality, which is what the grammar can account for must be distinguished from acceptability, the judgement made by native speakers.[22]

This argument is linked to the claim that English is not a finite state language and to the rejection of Markov's model as unable to isolate the set of all grammatical sentences.[23] Resuming Chomsky's stand taken in *Syntactic Structures*, Chomsky and Miller (1963) put forward detailed psychological and statistical arguments against the idea that grammar would be a Markov chain and that probabilities could be applied to syntactic structures, in particular because of the recursivity and discontinuity of natural language.

In addition it should be said that Chomsky's criticism essentially concerned the taxonomic view of corpus and discovery procedures, and not statistical methods in general. Actually, as far as they did not handle syntactic structures, Chomsky did not dispute the interest of statistics and probabilistic models in the study of language. Note that the following arguments made by Chomsky have not been mentioned by corpus linguists:

> Despite the undeniable interest and importance of semantic and statistical studies of language, they appear to have no direct relevance to the problem of determining or characterizing the set of grammatical utterances. (Chomsky, 1957: 17)

> Given the grammar of a language, one can study the use of the language statistically in various ways; and the development of probabilistic models for the use of language (as distinct from the syntactic structure of language) can be quite rewarding. Cf. B. Mandelbrot, "Structure formelle des textes et communication: deux études" *Word 10*.1-27 (1954); H. A. Simon, "On a class of skew distribution functions" Biometrika 42.425-40 (1955). (Chomsky, 1957 note 4: 17).

Miller and Chomsky (1963) agreed that Zipf's law as well as Mandelbrot's work, dealing with probabilities and word length in a text, had to be taken seriously, and their results discussed and verified: 'Miller and Newman (1958) have verified the prediction that the average frequency of words of length *i* is a reciprocal function of

---

[22] Later, Sidney Greenbaum (1976), one of the authors of *A Grammar of Contemporary English*, attempted to demonstrate that the acceptability of syntactic structures is influenced by their frequency of use.

[23] Named after Andrej A. Markov (1856–1922), who studied poetry as stochastic sequences of characters, a Markov chain is a sequence of random values whose probabilities at a certain time interval depends upon the value of the number at the previous time. Claude E. Shannon (1916–2001) used a Markov chain to create a probabilistic model of the sequences of letters in a piece of English text (Shannon, 1948). A Markov model of order *n* predicts that each letter occurs with a fixed probability, but that probability can depend on the previous *n* consecutive letters (*n*-gram). Since the 1950s Markov models have been used in Machine Translation and Natural Language Processing to disambiguate graphic units.

their average rank with respect to increasing length' (Miller and Chomsky, 1963: 461).[24]

Miller and Chomsky discussed Markov models more thoroughly in their paper, and agreed that, though they cannot be implemented on syntax to provide the set of grammatical sentences, they can be applied to lower-level production, such as phonemes, letters and syllables:

> Higher-order approximations to the statistical structure of English have been used to manipulate the apparent meaningfulness of letter and word sequences as a variable in psychological experiments. As k increases, the sequences of symbols take on a more familiar look and – although they remain nonsensical – the fact seems to be empirically established that they become easier to perceive and to remember correctly. ... We know that the sequences produced by k-limited Markov sources cannot converge on the set of grammatical utterances as k increases because there are many grammatical sentences that are never uttered and so could not be represented in any estimation of transitional probabilities. (Miller and Chomsky, 1963: 429)

Note that Kucera was concerned with Information Theory (Kucera, 1963) and used a Markov model in a comparative phonological study of Russian, Czech and German (Kucera and Monroe, 1968). He agreed with Chomsky's view that this type of model could only be applied to lower-level units and not to syntax and sentences.

To conclude this point, it has been shown that the Brown corpus could neither be a descendant of the taxonomy-oriented methods advocated by the post-Bloomfieldians nor the real target of Chomsky's criticisms. Concerning statistics and probabilities, Chomsky found certain types of statistical works quite valuable so far as they do not deal with syntax. Therefore, corpora like the Brown corpus, dedicated to word frequency counts, were not Chomsky's concern.

Conclusion

It has been shown that retrospective construction of a history aiming to legitimize Corpus Linguistics as an autonomous discipline, rests on a fair number of assertions and omissions. The Brown corpus has been presented as the key precursor by omitting pioneer works in Machine Translation or in computerized corpora in the area of dictionary making. Legitimization has been achieved by placing Corpus Linguistics at the heart of a revival of 1950s empiricism, in particular by making it a follower of the post-Bloomfieldian tradition. Still it was seen that the Brown corpus was submitted to

---

[24] George Kingsley Zipf (1902–1950) was a behaviorist linguist at Harvard and the founder of 'Dynamic Philology'. Zipf's law has been much used in statistical studies of vocabulary. Empirical data on word frequencies may by represented by an harmonic law: when the words of a text are ranked in order of decreasing frequency, the frequency of a word is inversely proportional to its rank (Zipf, 1949).
Benoît Mandelbrot (b. 1924), a French mathematician, developed a statistical model which provided a theoretical explanation for Zipf's law (Mandelbrot, 1954).

objectives quite different from any taxonomic machinery, in so far as its main concern was word frequency counts or statistical model testing.

Chomsky's arguments against the post-Bloomfieldians have been used to explain an alleged hiatus of corpus production which did not really occur. In fact, it can be assumed that there was no discontinuity between the present annotated corpora and vocabulary count corpora which were flourishing throughout the early twentieth century. On the other hand, Chomsky's criticism of corpora and statistical methods did not concern vocabulary counts. Rather, he seemed to find the use of Markov and word statistics models quite valuable, as far as they did not deal with syntax. Recall too that Kucera, one of the Brown corpus's authors, agreed with Chomsky on this point, using Markov's model in a comparative study of phonemes.

Connecting the first computerized corpora to the American tradition alone, in addition to being unfounded, has serious consequences. By ignoring the strong empiricist British filiation inherited from Firth's work, Corpus Linguistics has been deprived of a real precursor. Instead, two restrospective constructions were forged at the moment when NLP was technologically ready to invest in the field of corpora: a theoretical anti-precursor, one of the most famous theoretical linguist, i.e Chomsky; and a technical precursor, in fact a product, the Brown corpus.

It remains to be explained why the use of corpora, which has undeniably seen an unprecedented technological development and is valuable in most linguistic areas, absolutely needs be built up as an autonomous discipline. Another important issue is how to appraise the real impact of the increasing power of computers and the availability of linguistic data. Although corpus linguists invoke these technological developments as revolutionary for linguistic research, their real significance has as yet hardly been evaluated.[25]

References

Altenberg, Bengt. 1991. 'A Bibliography of Publications Relating to English Computer Corpora'. In: *English Computer Corpora: Selected Papers and Research Guide*. Ed. by Stig Johansson and Anna-Brita Stenstrom. Berlin: Mouton. 355–96.

Aarts, Jan & Willem Meijs. 1984. *Corpus Linguistics: recent developments in the use of Computer corpora in English Language Research*. Amsterdam: Rodopi.

Bar-Hillel, Yehoshua. 1960. 'The Present Status of Automatic Translation of Language' In: *Advances in Computers vol.1*. Ed. by Franz L. Alt. New York & London: Academic Press. 91–141.

Baayen, Harald. 2003. 'Probabilistic Approaches to Morphology' In: *Probabilistic Linguistics*. Ed. by Bod Rens, Jennifer Hay and Stefanie Jannedy. Cambridge & London: MIT Press. 229–287.

Bod, Rens, Jennifer Hay, Stefanie Jannedy (eds). 2003. *Probabilistic Linguistics*. Cambridge and London: The MIT Press.

---

[25] It seems a little difficult to follow Baayen (2003) when he claims that formal languages were developed in the 1960s instead of corpus investigations because at that time computers had computer power but no memory capacities. It seems exaggerated to claim that technological developments alone are responsible for theoretical orientations in language sciences.

Chomsky, Noam. 1956. 'Three models for the description of language'. *IRE Transactions on Information Theory*. IT-2: 113–124.

-----. 1957. *Syntactic Structures*. The Hague: Mouton.

-----. 1962. 'Transformational Approach to Syntax.' *Third Texas Conference on Problems of Linguistic Analysis in English May 9-12, 1958, Studies in American English*. Austin: The University of Texas. 124–158

Church, Kenneth and Robert L. Mercer. 1993. 'Introduction to the special Issue on Computational Linguistics Using Large Corpora'. *Computational Linguistics*. 19: 1–24.

Cori, Marcel and Jacqueline Léon. 2002. 'La constitution du TAL. Etude historique des dénominations et des concepts'. *Traitement Automatique des Langues* 43-3: 21–55.

Falk, Julia. 2003. 'Turn to the history of linguistics. Noam Chomsky and Charles Hockett in the 1960s'. *Historiographia Linguistica*. 30: 129–185.

Fillmore, Charles J. 1992. "Corpus linguistics' or 'Computer-aided armchair linguistics". In: *Directions in Corpus Linguistics. Proceedings of Nobel Symposium, 4-8 August 1991*. Ed. by Jan Svartvik. Berlin, New York: Mouton de Gruyter. 35 –60.

Francis, W. Nelson. 1998. 'A Pilgrim Progress: From philology to linguistics'. In: *First Person Singular III*. Ed. by E. F. K. Konrad. Amsterdam, Philadelphia: Benjamins. 59–70.

Garside, Roger, Geoffrey Leech, and Tony MacEnery. 1997. *Corpus Annotation, Linguistic Information from Computer Text Corpora*. London and New York: Longman.

Gellerstam, Martin. 1992. 'Modern Swedish Text Corpora'. In: *Directions in Corpus Linguistics. Proceedings of Nobel Symposium, 4-8 August 1991*. Ed. by Jan Svartvik. Berlin & New York: Mouton de Gruyter. 149–163.

Greenbaum, Sidney. 1976. 'Syntactic Frequency and Acceptability'. *Lingua* 40: 99–113.

Guiraud, Pierre. 1954. *Bibliographie critique de la statistique linguistique*. Utrecht & Anvers: Editions Spectrum.

Hockett, Charles F. 1957 [1948]. 'A note on structure'. In: *Readings in Linguistics I. The Development of Descriptive Linguistics in America 1925-56*. 3rd Edition. Ed. by Martin Joos. Chicago & London: University of Chicago Press. 279–280.

-----. 1968. *The State of the Art*. The Hague: Mouton.

Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics*. London & New York: Longman.

Kucera, Henry. 1963. 'Entropy, redundancy and functional load in Russian and Czech'. In: *American contribution to the fifth international congress of Slavists, Sofia*. The Hague: Mouton & Co. 291–318.

-----. 1998. 'A Linguistic Round Trip: From practice to theory and back'. In: *First Person Singular III*. Ed. by E. F. K. Koerner. Amsterdam, Philadelphia: Benjamins. 81–96.

Kucera, Henry and George K. Monroe. 1968. *A Comparative Quantitative Phonology of Russian, Czech and German*. New York: American Elsevier Publ.

Kucera, Henry and W. Nelson Francis. 1967. *Computational Analysis of Present Day American English*. Providence: Brown University Press.

Leech, Geoffrey. 1991. 'The State of the Art in Corpus Linguistics'. In: *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. Ed. by Aijmer Karin & Bengt Altenberg, London & New York: Longman. 8–29.

-----. 1992. Corpora and theories of linguistic performance. In: *Directions in Corpus Linguistics. Proceedings of Nobel Symposium, 4-8 August 1991* Ed. by Jan Svartvik. Berlin, New York: Mouton de Gruyter. 105–122.

-----. 2002. 'Geoffrey Leech' In: *Linguistics in Britain: personal histories*. Ed. by Keith Brown & Vivien Law. Oxford: Publications of the Philological Society: 155–169.

Léon, Jacqueline. 1999. 'La mécanisation du dictionnaire dans les premières expériences de traduction automatique (1948–1960)'. In: *History of Linguistics 1996 vol. II*. Ed. by David Cram, Andrew Linn and Elke Nowak. Amsterdam & Philadelphia: John Benjamins. 331–340.

Lexicologie et lexicographie françaises et romanes. Orientations et exigences actuelles. 12-16 nov. 1957. 1961. Paris: Ed. du CNRS.

McEnery, Tony and Andrew Wilson. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Mandelbrot, Benoît. 1954. Structure formelle des textes et communication. *Word* 10 (3): 1–27.

Manning, Christopher and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts & London: MIT Press.

Martin, Robert. 2000. 'Le Trésor de la langue française'. In: *Histoire de la Langue Française (1945–2000)* Ed. by Gerald Antoine & Bernard Cerquiglini. Paris: CNRS Editions. 969–979.

Miller, George A. and Noam Chomsky. 1963. 'Finitary Models of Language Users'. In: *Handbook of Mathematical Psychology Vol. II*. Ed. by R. Duncan Luce, Robert R. Bush and Eugene Galanter. New York: Wiley. 419–491.

Oostdijk, Nelleke and Pieter de Haan (eds.). 1994. *Corpus-based Research into Language*. Amsterdam, Atlanta: Rodopi.

Quirk, Randolph. 1960. 'Towards a description of English Usage'. *Transactions of the Philological Society*. 40–61.

-----. 2002. Randolph Quirk. In: *Linguistics in Britain: personal histories*. Ed. by Keith Brown and Vivien Law. Oxford: Publications of the Philological Society. 239–248.

Quirk, Randolph & David Crystal. 1966. 'On scales of contrast in connected English Speech'. In: *Memory of J.R. Firth* Ed. by Charles E. Bazell et al. London: Longman.

Quirk, Randolph, A.P. Duckworth, Jan Rusiecki, Jan Svartvik & A.Colin. 1964. 'Studies in the correspondence of prosodic to grammatical features'. In: *Proceedings of the Ninth International Congress of Linguists*. The Hague: Mouton: 679–691.

Quirk, Randolph & Joan Mulholland. 1964. 'Complex prepositions and related sequences'. *English Studies* 45: 64–73.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1972. *A grammar of Contemporary English.* London: Longman.

Quirk, Randolph & Jan Svartvik. 1966. *Investigating linguistic acceptability.* The Hague: Mouton.

Rosenblatt, Frank. 1958. The Perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review,* 65: 386–408.

Shannon, Claude E. 1948. 'A Mathematical Theory of Communication'. *The Bell System Technical Journal* 27: 379–423, 623–656.

Sinclair, John. 1991. *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Simpson Rita C. and John M. Swales (Eds.). 2001. *Corpus linguistics in North America: Selections from the 1999 Symposium.* Ann Arbor: University of Michigan Press.

Svartvik Jan (ed.). 1992. *Directions in Corpus Linguistics. Proceedings of Nobel Symposium, 4-8 August 1991.* Berlin & New York: Mouton de Gruyter.

Zipf, George Kingsley. 1949. *Human Behaviour and the Principle of Least-Effort.* Cambridge, Massachusetts: Addison-Wesley

Contact Details: jacqueline.leon@linguist.jussieu.fr

Jacqueline Léon
Université Paris 7
UMR7597, Histoire des Théories Linguistiques
case 7034
2, place Jussieu
75005 Paris
France

HENRY SWEET SOCIETY FOR THE HISTORY OF LINGUISTIC IDEAS

Abstracts from the 22nd Annual Colloquium

Darwin College, Cambridge; 18 March 2005

The Study of Child Language Acquisition in 19th-Century England

Paula Hellal and Marjorie Lorch
Birkbeck College, University of London

By the 1870s the developing mind of the child was beginning to be seen as a legitimate subject of scientific inquiry. The best-known works on the subject in England were Taine (translated from French in 1877) and Darwin (1877). The increasing interest in childhood was reflected in research undertaken to investigate the process of language acquisition. The young child's expressive language skills were an aspect of infant behaviour that could be systematically documented and analysed. This paper focuses on Darwin's 1877 account of his son's early language development.

   The motivation for Darwin's diary study was to investigate the normal pattern of child development. However, Darwin also believed, like other researchers of the period, that a careful, scientific investigation of the process might lead to insights regarding the origin of language in the human species. This paper will assess Darwin's ideas regarding language acquisition, both in the child and mankind, within the context of late 19th century theorising. English physicians, during this period, made little, if any, reference to these accounts of language acquisition in healthy individual children in their study of abnormal development. However, given the importance of Darwin as a scientific figure, his 1877 article was eventually noted by the medical community (e.g., Wyllie, 1894). This paper will conclude by reviewing Darwin's influence with respect to language development in present day research.

The History of Metaphor as a Feature of Linguistic Theories

Christine Horne
University of New Brunswick, Canada

This paper takes as its point of departure the assertion made by many contemporary linguists that metaphor is a ubiquitous and unavoidable feature of all natural languages. This very observation, however, brings into clear focus the virtual exclusion of metaphor from the dominant linguistic theories. Yet, has this always been the case? An examination of the place accorded to the study of metaphor in the history of linguistics sheds light on the changing scope of the field, and reveals the influence

of other disciplines on linguistic thought. Attitudes towards metaphor are contextualised in a comparative revisiting of Saussurean linguistics, I. A. Richards' views on metaphor as a necessary feature of a theory of language, and Chomskyan generative grammar.

Is There Progress in the History of Linguistic Ideas?

Werner Hüllen
University of Essen, Germany

The paper takes up a question which was raised during our meeting in Munich and which was discussed there in a controversial way.

As a test case, three semantic theories of the last century will be compared: (i) semantic field theory as explained first by J. Trier and later developed under the same term by structuralist linguists like Coseriu and Lutzeier, (ii) semantic marker theory as explained first by U. Weinreich und E. Nida and later developed under different terminology by Katz and Fodor and the followers of the early versions of the generative transformational standard theory, but also by structuralist representatives of semantics, and (iii) semantic model (or: schema) theory as first explained by the psychologist E. Rosch und linguists like J. Lakoff and R. Langacker.

All three theories are pertinent even today, if in forms and with terminologies which have changed since their origins. They will be compared and evaluated in the light of the question mentioned. It will be shown that all of them contribute (i) to semantics in the narrow sense of defining word meanings, and (ii) to wider interests which are combined with semantics. With reference to (i) they are rather traditional, with reference to (ii) they wish to be innovative. The question of progress (in the sense of linear improvement of linguistic knowledge in the course of time) has to be answered separately as concerning (i) and (ii). The general answer will be 'No, but...'.

Andrzej Gawronski:
Portrait of the Polish Linguist and Author of the Sanskrit Grammar

Iwona Milewska
Jagiellonian University, Krakow, Poland

Andrzej Gawroński was an outstanding and versatile scholar both of linguistics and indology. He was also a gifted translator. In his works he joined linguistics and the literature. The fruits of his short life include linguistic works on Sanskrit texts of different genres such as the classical drama Mrcchakatika, the novel Daśakumaracarita and the Buddhist works of the poet Aśvaghosha. He wrote also a genuine dissertation on the origins of Indian theatre. His grammar of Sanskrit, edited posthumously in 1932, has since been used by all Polish adepts of Sanskrit. As well as his works on linguistics, he also prepared several translations from Sanskrit and Persian into Polish. These translations are remarkable for their beauty. He was a man of genius, extra-

ordinarily talented and gifted not only for languages. Parts of his works were written in German, English or French but the vast part is left only in Polish.

Shelf-Life, Use-By Dates and Time Horizons
in the Historiography of Linguistics

David Cram
Jesus College, Oxford

The aim of this programmatic paper is: (i) To outline a framework for linguistic historiography which finds its underpinnings within linguistics itself, and more specifically within a pragmatic theory of temporal deixis; (ii) To establish a connection between metahistoriographical issues in historiography and explanatory issues in current linguistics.

The approach is based on what has been called the 'specious present' (as distinct from the 'punctual' present). "The practically cognised present is no knife-edge, but a saddle-back, with a certain breadth of its own on which we sit perched, and from which we look in two directions into time. The unit of composition of our perception of time is a *duration*, with a bow and a stern as it were – a rear-ward and a forward-looking end" (William James 1890). This will be exemplified in our case in terms of the 'shelf-life' of publications and references. The linguistic 'now' is centered on the cutting edge of current research, but also has a duration delimited by time horizons (to mix metaphors).

It will be argued that the history of ideas involves a sequence of 'nows', each with its own set of time horizons. Linguistic historiography involves reconstruction of the 'nows' of the past. Conversely, evaluation of contemporary theories involves the 'history of the present'.

PUBLICATIONS RECEIVED

(to 1st May 2005)

Members of the Society have been kind enough to donate the following publications to the HSS Library. Further contributions, which are very welcome, should be sent to:

> Dr Richard Steadman-Jones
> Dept of English Language & Linguistics
> University of Sheffield
> Sheffield   S10 2TN

Monographs by individual authors will be reviewed wherever possible; articles in collected volumes will be listed separately below, but, like offprints and articles in journals, will not normally be reviewed. It would be appreciated if the source of articles could be noted where not already stated on the offprints.

The Society is also very grateful to those publishers who have been good enough to send books for review.

BOOKS AND PAMPHLETS

FORMIGARI, Lia
A History of Language Philosophies.
Amsterdam, Philadelphia: John Benjamins (Amsterdam Studies in the Theory and History of Linguistic Science), 2004. x, 252 pp. ISBN • 90-272-4596-7 (Eur.); ISBN • 1-58811-561-5 (US). EUR 44,00; US$ 52.95.

HÜLLEN, Werner
A History of Roget's Thesaurus: Origins, Development, and Design.
Oxford, New York: Oxford University Press, 2004. ix, 410 pp. ISBN • 0-19-925472-9 • 0-19-928199-8 (paperback). £86.50; £22.95 (paperback).
(Professor Hüllen's book is dedicated 'To my friends in the Henry Sweet Society'.)

NORDEGRAAF, Jan
Een kwestie van tijd: Vakhistorische studies.
Münster: Nodus Publikationen, 2005. 191 pp. ISBN • 3-89323-291-5. EUR 35.50.

SCHULTINK, H. (PORTIELJE, Cecile A. and NOORDEGRAAF, Jan (eds))
Van Onze Taalkundige Mederwerker: Kronieken 1954-1962.
Amsterdam: Stichting Neerlandistiek VU Amsterdam, Münster: Nodus Publikationen (Cahiers voor Taalkunde 22), 2005. x, 214 pp. ISBN • 90-72365-86-0 (Stichting Neerlandistiek VU Amsterdam.); ISBN • 3-89323-528 (Nodus). EUR 35,50.

PERIODICALS

Historiografia da Lingüistica Brasileira, Boletim VII (2004).

NEWS AND ANNOUNCEMENTS

Stylesheet for Submissions
to the Bulletin of
the Henry Sweet Society for the History of Linguistic Ideas

SUBMISSION PROCEDURE: All submissions should be sent to the editor in electronic
format, preferably as an e-mail attachment in .rtf format (if this is not possible then
please use .doc). If special characters, including IPA or characters with umlauts for
instance, are used the author is also requested to send a hard copy, or a pdf version to
the editor.

PEER REVIEWS: All submissions are read by the editor and passed on to a suitable
reviewer on the Henry Sweet Society Committee, or when this is not possible the
editor will approach someone outside the committee with special knowledge in the
relevant area.
        After the peer review, comments are sent to the author together with a note to
tell him/her if their submission has been accepted and if so if any further work needs
to be done before it can be published. (This is normally done via e-mail, so the person
submitting something for the Bulletin is asked to make sure that he/she includes an e-
mail address that is checked on a regular basis.)

ABSTRACTS: Authors of (short) articles are also asked to submit a 150-200 word
abstract which can be used to give information of the contents of the most current
issue of the Bulletin on the Henry Sweet Society Web pages. In the future this might
also be published as part of the article, in order to help readers get a quick overview of
the contents.

PROOFS: Before publication the first proofs are sent to the author and after they have
been approved by the author the paper is sent to a proof reader. If further changes need
to be made according to the proof reader a new set of proofs will then be sent to the
author.

TITLE: Centred, bold, 16 points. The first letter of each content word should be upper
case and the rest of the main title should be in lower case. If there is a subheading this
should be presented in italics, 14 points, without capitalisation of the first letter of
content words.

HEADINGS FOR REVIEWS: The heading should be left-aligned, 16 points, Times New
Roman. The author(s) / editor(s) of the reviewed volume should be presented in bold,
on the next line the title should be given in bold italics, followed by place of
publication, publisher, year, pages and the price (when known) on the third line in
regular, 14 points. If a discount is available for members of the Henry Sweet Society
for the History of Linguistic Ideas it would be good if information of this could also be

included. The name of the reviewer should appear underneath the information about the publication, leaving one blank line between (bold, 13points, left aligned, preceded by 'Reviewed by:').

Example:
David Cram, Jeffrey L. Foreng and Dorothy Johnston (eds.)
Francis Willughby's Book of Games. A Seventeenth-Century Treatise on Sports, Games and Pastimes.
London: Ashgate, 2003. 344 pp.
£65, USD 114.95 (25% discount for members of the Henry Sweet Society)

Reviewed by: Werner Hüllen, Düsseldorf.

AUTHOR: The name, possible affiliation and place of residence of the author should be given below the heading of articles, reports, proposals etc. leaving two lines between the heading and the name. This should be 13 points, centred, the author's name given in bold and the affiliation, and/or place of residence given in regular font style on the next line.

        Authors of reviews are also asked to include their name and affiliation after the review heading. One line should be left between the review heading and the author. The author's name should be left aligned, bold, 13 points, Times New Roman and it should be introduced by '**Reviewed by:**' and the name followed by a comma and then the reviewer's affiliation (and/or place of residence) in regular, 13 points, Times New Roman (see above).

        The author's address details should be given at the end of the article. These should be aligned to the left-hand margin of the page. Two lines should be left between the references and the details about the authors. These details should come after the heading 'Contact Details' (13 points, left aligned, bold) and should include postal address and e-mail address, unless the author would prefer not to publish one of those and then this should be discussed with the editor.

Example:
Contact details: a.m.t.tiedemann@rug.nl [tab: 3.75 cms]

BODY TEXT: The body of the text should be 13 points and the alignment should be justified. The first paragraph should begin with a drop cap that stretches over 2 lines. All other paragraphs should begin with a 1.27-cm indentation of the first line, except the first line of a new section which should not be indented.

SECTIONS OF THE BODY TEXT: It is up to the author if he/she would like to divide their paper into sections. Sections should be divided into 1., 2., 2.1, 2.1.1. The headings for these sections should be in Bold Italic 13 points for 1., 2. etc., italics only for 1.1, 1.2, etc. and underlined for 1.1.1, 1.1.2, etc. There should always be two blank lines before a 1., 2. etc. section starts, one blank line before a 1.1, 2.1, etc. section start. After each heading there should also be one blank line.

QUOTATIONS: Quotations should be clearly marked as such, with the reference given in the following manner (Smith, 1999: 34), unless the author is mentioned by name in the same sentence then (1999: 34) should be used. Quotations that are less than three lines long should be cited within citation marks in the text ['x', "x"]. Authors are requested to use curved quotation marks and to be consistent in their use of single or double. Quotations more than three lines long should be indented by 1.27 cm on the left-hand side, and should **not** be surrounded by quotation marks. If the author chooses to leave out part of a quotation this should be represented by [...].

REFERENCES: All works that are referred to in the paper should be given full references at the end of the paper. Full names (given name and surname) should be given for all authors, unless an author is referred to who prefers to use initials only in their own publications, then initials should normally be used. However, if in such cases the full names are in general circulation it is up to the author of the submitted text to decide whether to represent that scholar by his/her full name or only by initials and their surname. References should be given in the following style:

Monographs

Max Müller, Friedrich. 1862. *Lectures on the Science of Language Delivered at the Royal Institution of Great Britain in April, May and June, 1861.* New York: Charles Scribner.

-----. 1865. *Lectures on the Science of Language Delivered at the Royal Institution of Great Britain in February, March, April and May, 1863.* New York: Charles Scribner.

Article in journal

Hancher, Michael. 1981. 'Humpty Dumpty and Verbal Meaning.' *Journal of Aesthetics and Art Criticism.* 40: 49–58.

Rastall, Paul. 2001. 'Richard Chenevix Trench – not just a populariser?' *Bulletin of the Henry Sweet Society for the History of Linguistic Ideas.* 37: 22–39.

Sutcliffe, Patricia Casey. 2001. 'Humboldt's *Ergon* and *Energeia* in Friedrich Max Müller's and William Dwight Whitney's Theories of Language.' *Logos and Language (Topics in the Historiography of Language Theory).* 2 (2): 21–35.

Reprint or Paper in an edited volume

Humboldt, Wilhelm von. 1836 [1963]. *Ueber die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts.* In: *Werke in fünf Bänden.* Vol. III Schriften zur Sprachphilosophie. Ed. by Andreas Flitner & Klaus Giel. Darmstadt: Wissenschaftliche Buchgesellschaft. 368–756.

Rydén, Mats. 1999. 'Axel Erdmann: Sweden's First Professor of English.' In: *Thinking English Grammar: to Honour Xavier Dekeyser, Professor Emeritus.* Ed. by Guy A. J. Tops, Betty Devriendt & Steven Geukens. Leuven: Peeters; Hadleigh: BRAD. 297–305.

Reprint edition
Whitney, William Dwight. 1873 [1987]. *Oriental and Linguistic Studies*. Vol. 1.
    Delhi: Sri Satguru Publications.

Encyclopedia entry
Söhnen-Thieme, Renate. 1994. 'Müller, Friedrich Max.' *The Encyclopedia of
    Language and Linguistics*. Oxford & New York: Pergamon Press. 2617–2618.

LIFE DATES: All scholars that are mentioned in the Bulletin should normally be
provided with information about when they lived on their first mention in an article or
review. However, scholars that are still alive need not be given life dates, unless the
article / review treats the recent history of linguistics and the author feels that it may be
useful to include information about when the scholar was born, an alternative in this
case may be to include information about when they graduated.

Therese Lindström Tiedemann, *Groningen*
Editor
November 2004, revised June 2005

Vorläufiges Programm
First East Asian SGdS-Conference on the History of Linguistics

28.-30.10.2005, University of Hong Kong (Hongkong, VR China)
Organisation: Christopher Hutton, Peter Schmitter und Hans-Georg Wolf

I. Zeitrahmen
Anreise: Donnerstag, den 27.10.2005
Konferenztage: Freitag, den 28.10.05, bis Sonntag, den 30.10.2005
Abreise: Montag, den 31.10.2005

II. Vorträge (in alphabetischer Reihenfolge)

Kuk-Hyun **Cho** (Hanyang University, Seoul, Korea):
Rezeptionsgeschichte der Textlinguistik in Korea. Hauptinteresse und
Anwendungsbereiche

Xiao Ling **Cui** (Beijing Foreign Studies University, China):
40 Years' Development of Chomsky's Linguistic Theory in Korea

Henryk **Duda** (Tokyo University of Foreign Studies, Japan):
Memento Sexus! Jan N. Baudouin de Courtenay's Remarks on Grammatical Gender in
the Light of Recent Linguistic Theories

Hiroyuki **Eto** (Nagano College of Nursing, Japan):
Philological and Exegetical Approach into Language and Culture in the History of
Language Study in Japan

Camiel **Hamans** (European Parliament, Brussels/Strasbourg, Belgien/Frankreich):
The Minority Language Debate: The Case of Yiddish in the Dutch Language
Landscape

Christopher **Hutton** (The University of Hong Kong, China):
Diversity, Hybridity and Nationalism in Late Nineteenth Century Linguistics

Jae-Won **Lee** (Duksung Women's University, Seoul, Korea):
Zur Geschichte der Analyse kommerzieller Werbung in Korea

Markus **Meßling** (Freie Universität Berlin, Deutschland):
Wilhelm von Humboldt and the "Orient". On Edward W. Said's Remarks on
Humboldt's Orientalist Studies

Yo-Song **Park** (Cheju National University, Jeju, Korea):
Zur Vorgeschichte der Semiotik. Koreanisches Pansori *Heungboga* als eine
semiotische Konzeption

Jamin **Pelkey** (La Trobe University, Australia):
A Brief History of Yi Linguistics: Dialectic, Paradox and Process

Tommaso **Pellin** (Ca'Foscari University of Venezia, Italien):
The Coinage of Chinese Grammatical Lexicon: The Influence of the Historical
Context

Peter **Schmitter** (Hankuk University of Foreign Studies, Seoul, Korea & Westfälische
Wilhelms-Universität Münster, Deutschland):
Von der 'naturgegebenen Richtigkeit' zur 'Arbitrarität'. Zum historischen Hintergrund
der Saussure'schen These "le signe linguistique est arbitraire"

Hyung-Uk **Shin** (Hankuk University of Foreign Studies, Seoul, Korea):
Deutschunterricht in Korea: Entwicklungstendenzen und die Herausbildung des
Faches 'Deutsch als Fremdsprache'

Hiroyuki **Takada** (Gakushuin University, Tokyo, Japan)
Zur Bedeutung der Wortbildung als Mittel der Sprachkultivierung bei G.W. Leibniz

Manabu **Watanabe** (Gakushuin University, Tokyo, Japan)
Ein historisch-historiographischer Überblick über die Sprachkultur in Japan von der
Edo-Zeit bis hin zur Gegenwart

Wolfgang **Wildgen** (Universität Bremen, Deutschland):
Interkulturelle Semantik und künstliches Gedächtnis. Ein Vergleich der Beiträge von
Giordano Bruno (1548-1600) in Europa und Matteo Ricci (1552-1610) in China

Hans-Georg **Wolf** (The University of Hong Kong, China):
British Educational and Linguistic Policy in the Trusteeship Territories

Sheng-shiung **Wu** / Chia-lu **Chiang** (Taiwan)
Phonological Study on Sino-Japanese as recorded in Wamyo ruijishyo

Weitere Auskünfte erteilen:
Dr. Christopher Hutton (chutton@hkucc.hku.hk)
Prof. Dr. Peter Schmitter (schmipe@uni-muenster.de)
Dr. Hans-Georg Wolf (hanswolf@hkucc.hku.hk)

The Second Vivien Law Prize in the History of Linguistic Ideas

In memory of Dr Vivien Law (1954-2002), and thanks to her generosity, a prize has been established by the Henry Sweet Society for the best essay submitted on any topic within the history of linguistics.

The competition is open to all currently registered students, and to scholars who have received their PhD or equivalent qualification within the last five years. Members of the Executive Committee of the Society may not apply. Applications from non-members are welcome.

The prize consists of £100 and publication of the winning essay in the *Henry Sweet Society Bulletin*. Others of the essays submitted may also be published where appropriate. The prize will not be awarded if none of the submitted essays is deemed to be worthy of publication. The prize-winner is also entitled to one year's free membership of the Society and will receive a free copy of Vivien Law's *The History of Linguistics in Europe* from Cambridge University Press.

The prize will be awarded by the Executive Committee on the recommendation of a Prize Committee drawn from its members. The committee will be looking for an exciting and original approach to the history of linguistics, either in the choice of topic or in the way it is treated, and for the highest standards of research and presentation. The essay should not have been previously published.

The closing date for submissions is 30 September 2005. Entries may be written in English, French or German, and should follow the style-sheet for the *Henry Sweet Society Bulletin*. They should not exceed 8000 words, including references, footnotes, tables, appendices, etc. Four hard copies of the essay, and one in electronic form, should be sent to the Chairman of the Executive Committee (Dr David Cram, Jesus College, Oxford OX1 3DW), by the closing date. The Committee's decision will be final. The winning entry will be announced in the May edition of the Bulletin, but all entrants will receive notification of the outcome by the end of December.

Vivien Law studied Classics and German at McGill University, Montreal, before pursuing PhD studies at Cambridge. She was successively a Fellow at Jesus, Sidney Sussex and Trinity Colleges in Cambridge, and held the only lectureship in the world dedicated to the history of linguistic thought (in the Cambridge Department of Linguistics). In the late 1990s she was made Reader in the History of Linguistic Thought and a Fellow of the British Academy. Her academic interests were wide-ranging, but she was associated above all with her work on medieval grammars.

Worldwide Universities Network

Language Reform Project

In October 2004 we were pleased to announce that the Worldwide Universities Network (http://www.wun.ac.uk) – a consortium of research-intensive universities from Europe, USA and China – had embraced the History of Linguistics as one of the humanities disciplines it wished to support. We agreed at the outset that the study of language reform activities, past and present, would provide a focus for our work. The history of language reform is a research area in which colleagues from across the WUN network are particularly actively involved, but it is clear that involvement is not going to be limited to researchers based in these partner universities.

Following a period of consultation we have now identified a number of specific short-term goals for the language reform project.

1   Leverhulme Trust International Network
The Leverhulme Trust (http://www.leverhulme.org.uk) is a charitable foundation, which funds research via a range of different schemes, one of which is entitled *Academic Collaboration: International Networks*.   Andrew Linn (Sheffield) is submitting a bid to this scheme to set up a formal network of scholars working on the issue of language reform in its various guises. Funding covers the salary of a network coordinator for up to two years, the cost of organising workshops / seminars / conferences, and the travel and accommodation costs incurred by members of the network in taking part in these events. Up to 6 overseas partners may be included, and it is likely that these 'overseas partners' will be groups of researchers in the USA, Germany, Norway, France, and possibly Canada and The Netherlands. There is no closing date for the application, and the application process has two stages.

2   Prescriptivism on-line
In collaboration with the Supercomputing Applications Centre at the University of Illinois Urbana-Champaign, Doug Kibbee is mounting an on-line searchable database of prescriptive statements concerning French in the period 1800-2000. Computer scientists at the University of Illinois are helping him mount this into a collaborative web-site, so that scholars involved in the WUN project can add similar materials to the same site. The software will be distributed and installed for participants who come to the ICHoLS conference at UIUC in September.

We are proposing extending this initiative to include similar electronic collections of prescriptions concerning other languages across a similar historical timespan. The ultimate goal would be to have these interlinked, such that it would be possible to search for views on specific topics across a range of languages in order to ask questions about the universality or otherwise of prescriptive judgments and their

motivations. Some databases already exist and others will need to be set up, and it is expected that this project will initially develop piecemeal via smaller local grants, enabling linguists to work with specialists in humanities computing.

3   Publications series

We are producing a series of publications relating to language reform with the international publisher, Continuum Books (http://www.continuumbooks.com). Andrew Linn is finalising the proposal with Continuum's linguistics editor, Jenny Lovel, and there is plenty of enthusiasm from the publisher's side. The first collection of publications will be four volumes charting key debates on specific language reform topics from across history. The volumes will address: 1) artificial languages; 2) prescription and standards; 3) spelling and orthographic reform; 4) regulation and intervention. Because of the publisher's market, there is likely to be a strong English-language bias in these volumes. We have also discussed a separate series of monographs on language reform topics, and these could address reform issues relating to any language or languages.

4   Conference in Bellaggio, Italy

Doug Kibbee, Joseph Subbiondo (California Institute for Integral Studies) and Mark Amsler (University of Wisconsin – Milwaukee) are preparing a proposal to the Rockefeller Foundation, to support a conference tentatively titled "Perfecting Language". If funded, the invited participants will gather in Bellaggio to discuss language reform efforts and their success (or failure) through history. This conference would take place in 2006 or 2007, according to the availability of the conference venue in northern Italy.

We hope that you feel these are all positive and exciting proposals, which will serve to bring together the community of scholars in and beyond the WUN universities, and engender a better understanding of the nature of language reform in the past to inform language reform activities of the future. Please feel free to get back to us with comments or suggestions, and we will keep HSS members informed as things progress.

Andrew Linn (A.R.Linn@shef.ac.uk)
Doug Kibbee (dkibbee@uiuc.edu)

The Henry Sweet Society for the History of Linguistic Ideas

Subscription rates

Ordinary Members: £15 (£14 if paying by standing order, £16 if paying by credit card on PayPal)
Associate Members (within three years of graduation only): £5 (£4 if paying by standing order, £6 if paying by credit card on PayPal)

UK members are reminded that subscriptions may be set against Income Tax. The tax reference is: H.O. Ref. T 164418711186/MT.

Ways to pay

1. Paying by standing order through a British bank (reduced rate £14 or £4)
To arrange to pay your subscription annually by Standing Order, contact the Honorary Treasurer and ask for the appropriate form to be sent to you.

*2. Paying by Pay-Pal (£16 or £6 if using a credit card; £15 or £5 otherwise)**
Please follow the instructions at the PayPal website, http://www.paypal.com, to send your subscription to nicola.mclelland@nottingham.ac.uk. If you have not used PayPal before, you will need to register as a PayPal user before you can make a payment. This takes about ten minutes to set up, and you will need to give details of a credit card or other bank card.
In the "Notes" section of your message, state that the payment is a subscription for the Henry Sweet Society, for which year(s), and include your full name and full contact details. (Under "Payment type", select "Service" - if you select "Quasi-Cash", your bank may charge interest, as if it were a cash withdrawal on your credit card).

*Using a credit card is the easiest way to use PayPal, but unfortunately the Society is then charged a fee to draw down the funds, so we must ask you to pay a higher rate.

3. Paying directly into our Dutch bank account (24 Euros / 8 Euros):
Members in The Netherlands may pay subscriptions directly into the Society's bank account, Girorekening 8121692, with the Netherlands POSTBANK, Postbus 94780, 1090 GT Amsterdam, International Banking account number IBAN/NL/89PSTB000, BIC/Swift code PSTBNL21.

4. Paying directly into our British bank account (£15 or £5)
Alliance & Leicester, Account number 56 533 4204, IBAN GB 78 GIRB 7250 0565 3342 04, BIC/Swift code GIRBGB22. Please notify the treasurer separately that you are paying directly into the account.

5. Cheque or sterling draft sent direct to the treasurer (£15 or £5)
Send your cheque or sterling draft for £15 or £5 to the treasurer at the address below.

6. For members in the USA and other members paying in US dollars ($30 or $15)
Members in the USA and others paying in US dollars should send their dues to
Professor Joseph L. Subbiondo, President, California Institute of Integral Studies,
1453 Mission Street, San Francisco CA 94103, U.S.A. (E-mail: josephs@ciis.edu)

Dr Nicola Mclelland (nicola.mclelland@nottingham.ac.uk)
Honorary Treasurer
Department of German
University of Nottingham
NG8 1FN, England