

A3

The separation between grammar and lexicon is a paradigm case of accommodating vested interests: of formalist linguists who want to equate all of linguistics with ‘grammar’ or ‘syntax’; of teachers who prepare one tidy lesson on ‘grammar’ and another on ‘vocabulary’; and of authors and publishers who produce and market ‘grammars’ and ‘dictionaries’ as separate enterprises. The interests have been most complex and subtle in linguistics, where the lexicon has gotten blamed for defying the methods of formalist description. Already for Sweet (1964 [1899]: 73), grammar consisted of ‘general laws’, and the lexicon of ‘isolated facts’; for Saussure (1966 [1916]: 133) ‘grammatical’ languages have the highest ‘motivation’ whereas ‘lexicological’ languages are the most ‘arbitrary’; for Chomsky (1965: 86f), an ‘advantage’ of ‘the lexical entries’ should absorb all the ‘idiosyncrasies’ and ‘irregularities of the language’. Such views indicate why ‘linguistic theory’ has shown so little interest in lexicology as compared to ‘grammar’. If accepted the lexicon might just be made into a handy limbo to place any ‘irregularities’ that didn’t fit properly into your ‘grammar’.

The major counterpoint has been the unified concept of the lexicogrammar regarded in systemic functional linguistics as ‘the inner core of language’ (Halliday 1994a: 15). Evidence for the functional unity of its two sides can be found in several sources. The same or similar meanings or functions may be ‘lexicalised’ in one language and ‘grammaticalised’ in another; many lexical items entail grammatical constraints, while many grammatical patterns prefer certain types of lexical items, as large-corpus work has made obvious. And when infants shift from spontaneously designed sounds over to real lexical

25 items, their grammar soon begins assimilating to the mother language (Painter
1984). Still, the two sides clearly differ in their evolution; the lexicon changes
faster, accepts more deliberate innovations, and forms less systematic classes
than does the grammar; and the lexicon has a far more diverse and multiple
range of functional orders, reflecting the normally improvised ways in which
30 cultures use vocabulary to express and classify particular objects, events, actions,
and so forth. A vital and still very open question is whether ballooning effects
might occur similar to those noted above at the lexical end but more toward the
grammatical end, where they could be much harder to identify and compensate.
The unity of lexicogrammar as a principle of functionalist linguistics predicts
35 they would, especially when a corpus has heavy proportions of mass media
discourse, which, among other things, likes to grammaticalise female gender, as
we shall soon see. Only very large queries for colligation types can enlighten us
here.

40 Ironically, English — the very language for which the most formal grammars
have been devised — lexicalises far more functions than do many languages.
This trend is still under way as the language evolves over historical time, e.g.,
to take over former functions of forming diminutives with ‘-let’, which is
hardly productive except for coinages by analogy to ‘bracelet’ such as ‘anklet’,
45 and feminines with ‘-ess’, which is politically questionable but still found in
a few items dear to mass media discourse, such as ‘heiress’, ‘millionairess’,
‘adventuress’, ‘temptress’, ‘enchantress’, ‘seductress’, and of course ‘murderess.’³
So a unified functional lexicogrammar should be particularly useful for
English; and putting a large corpus ‘into’ or ‘behind’ it may well be the major
50 prospect for significant progress in linguistics toward coverage, convergence,

and consensus in the coming decades. We can expect to find many delicate and supportive interactions between ‘grammaticalising’ and ‘lexicalising’, yielding a characteristic profile for each language, language variety, register, discourse domain, and so on.

55

Fitting the corpus to the lexicogrammar will be a daunting challenge. Theory-driven access would proceed from an already formulated lexicogrammar like Halliday’s (1994a) and flesh it out with the data; practice-driven access would proceed from the corpus and gradually build up the lexicogrammar that the data
60 seem to recommend or require. The most plausible prospect is a convergence of these two strategies, whereby a Hallidayan lexicogrammar will be both applied and revised for large sets of corpus data (Beaugrande 1997d), with human analysts operating upon supports from steadily more sophisticated ‘functional’ software. The main advantage is that a Hallidayan lexicogrammar is linguistic,
65 cognitive, and social all at once, whence its three ‘meta-functions’: the ‘textual’ being what language gets used and why, the ‘ideational’ being what gets talked about and how, and the ‘interpersonal’ being who talks to whom (cf. Halliday 1967-68, 1985a, 1985b, 1994a). Also, systemic functional linguistics has long
70 been seeking to build the criteria of selection and combination directly into the description of the repertoires of options by means of ‘networks’ that can be used for computational implementations as well (e.g. Matthiessen and Bateman 1991).

The Hallidayan lexicogrammar is also emphatically ‘semantic’ and ‘functional’ in its categorisation of ‘processes’, such as ‘mental’ versus ‘behavioural’, now
75 graphically assembled in a circular multicoloured disk on the cover of Halliday’s revised Introduction of 1994. Based on my own practical analyses of data, I

have suggested several changes in the design and terminology — e.g., enriching the ‘aspect’ system, or having the more general term ‘semiotic processes’ in place of his ‘verbal’; or ‘cognitive’ and enactive processes’ in place of his ‘mental and ‘behavioural’, which might carry overtones of Cartesian dualism. But the ‘semantic’ grounding is surely fundamental to any functionalist lexicogrammar that could be powerfully interfaced with large-corpus data.

How ‘semantic’ and how ‘grammatical’ it needs to be are open questions for which corpus research might finally provide some data-driven answers. Halliday has cautioned that ‘all categories employed must be clearly “there” in the grammar of the language’, ‘not set up simply to label differences in meaning’; without some ‘lexicogrammatical reflex’, such ‘differences’ are not ‘systemically distinct in the grammar’ (1985a: xx). The toughest problem I see there is that in some languages, notably English, numerous ‘reflexes’ are formally systematic or distinctive only in certain sectors. In the English verb system, for example, the differences among process-types are most formally distinctive in the unmarked imperatives, presumably following the cognitive and social constraint that you normally only command an action for a genuine intentional agent capable of exerting the effort to perform and control it, e.g., for Enactive Processes like ‘sit down!’ or ‘don’t go away!’ more than for Cognitive ones like ‘?know the answer!’ or ‘?don’t understand this message!’ (see Beaugrande 1997a: ch. IV for the whole system). We need to work out from extensive corpus data how far such formal distinctions are in fact reliable or how far they can be offset by specific lexical counter-constraints, e.g., for a Cognitive like ‘think about the answer!’ or ‘don’t misunderstand this message!’