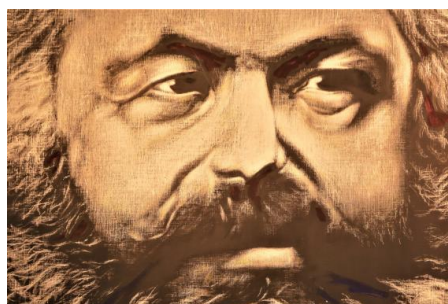

KARL MARX: A MANIFESTO FOR CORPUS LINGUISTICS

Karl Marx was born in Trier in 1818. In 2018 he celebrated his milestone birthday.

He was kind enough to take time off from the many parties and exhibitions to give a talk to the local English Linguistics Circle, and to place these notes for his talk in the University archives.



A spectre is haunting linguistics – the spectre of big data. All the powers of old linguistics have entered into an unholy alliance to exorcise this spectre. Was it not Chomsky – who ought to know better – who said:

“The sentence *I live in New York* is fundamentally more likely than *I live in Trier, Rheinland-Pfalz.*”

But surely this is parody? And where is the theory in opposition that has not been parodied by its opponents in power? Where is the opposition that has not hurled back the parodic reproach against its reactionary adversaries? Do they not see that quantitative differences, beyond a certain point, pass into qualitative changes?

It is high time that Corpus Linguists should openly, in the face of the whole world, publish their views, their aims, their tendencies, and meet this nursery tale of the spectre of big data with a manifesto of its own. To this end, Corpus Linguists of various nationalities have assembled in Trier and sketched the following manifesto, to be published in English, French, German, Esperanto, Xhosa, Lowland Scots, Letzebürgesch and Trierer Platt.

*

Voloshinov has taught us of one unholy alliance, the entanglement of religion and linguistics:

“The first philologists and the first linguists were always and everywhere priests. History knows no nation whose sacred writings or oral tradition were not to some degree in a language foreign and incomprehensible to the profane. To decipher the mystery of sacred words was the task meant to be carried out by the priest-philologists.”

The history of all hitherto existing linguistics is the history of empirical struggles, patrician and plebeian, posh accents and low-prestige dialects, literary languages of high culture and unwritten vernaculars of the proletariat, the classical Sanskrit of the Vedic scriptures and the colloquial usage of the masses, in a word,

oppressor and oppressed, who have stood in constant opposition to one another, and carried on an uninterrupted, now hidden, now open fight, a fight that each time ended, either in a revolutionary reconstitution of theory, or in the total confusion of poor students trying to make sense of the latest textbook.

We see, therefore, how modern linguistics is itself the product of a long course of development, of a series of revolutions in the modes of alienation, corroboration, distortion, exemplification, generalization, misrepresentation, production, refutation, simplification, verification. Let us recapitulate in general outline the phases that linguistics went through from the 19th to the 21st century.

Three main periods are unmistakable. Saussure stated the main principles of structuralist idealism. He provided the thesis and the antithesis – *langue* and *parole* – but not the synthesis. (And he died a very unhappy bunny.) Bloomfield recognized the real behaviour of living men and women. He studied the materialist basis of the languages of the native peoples of North America, but then told behaviourist fairy tales about Jack fetching apples for Jill. (Yet he seemed unaware of his own slide into abstraction.) Chomsky reinstated thesis and antithesis in an even more rigid form – *competence* and *performance* – again with no synthesis. (But he must be praised for turning from linguistics to politics.)

Hegel remarks somewhere (but I admit that I can't for the life of me find the reference) that all great world-historic facts and personages appear, so to speak, twice. He should have continued: not just twice, but in a constant oscillation between rationalist and empiricist, between tragedy and farce.

All science would be superfluous if appearance and essence directly coincided. So, we might add: just as rationalism failed, just as empiricism was suffocated by a fairy tale, just as rationalism again failed, finally we must look to the synthesis offered by Kant. Rationalism offers form without content, empiricism offers content without form. In Kant's own words:

“Thought without content is empty, perception without concepts is blind.”

Men and women make their own history, but they do not make it as they please; they do not make it under self-selected circumstances, but under already existing circumstances, given and transmitted from the past, sometimes by prescriptive teachers in the schools, who think we shouldn't end sentences with a preposition. The traditional grammars of all dead generations weigh like a nightmare on the corpus-based grammars of the living.

Language is practical consciousness that exists also for other speakers, and for that reason alone it really exists for me personally. Linguistics is theoretical consciousness. But whereas some linguists have only interpreted the world, others have tried to change it, by their study of the language of politics, laws, morality, and religion. Some “critical discourse analysts”, in their studies of language and power, have claimed – with justification – that consciousness is the first step towards emancipation.¹

¹ Note from the organizers. Marx did not give a reference here, but he was probably thinking of Norman Fairclough's 1989 book *Language and Power*.

In direct contrast to German philosophy which descends from heaven to earth, here we ascend from earth to heaven. As Jonathan St Clair has said:

“Language looks different when you look at a lot of speakers at once.”

It is indeed not only the language of men and women that determines their being, but their understanding of language in its social domain that determines their consciousness.

But the “critical discourse analysts” have taken my name in vain if they fail to recognise that all this requires appropriate methods of analysis. There is no royal road to Corpus Linguistics, and only those who do not dread the fatiguing climb of its steep paths – as they master Perl, R, multivariate analysis (not to mention the more obscure syntax of regular expressions and the convoluted pattern matching conventions of the latest concordance software) – have a chance of gaining its luminous summits.

The consequences of big data may be summed up in one sentence: Abolish all introspective data. Introspection is the opium of the masses. Introspection is the impotence of the human mind to deal with quantitative techniques and statistical procedures which it cannot understand.

Corpus Linguists disdain to conceal their views and aims. They openly declare that their ends can be attained only by swamping their opponents with so much corpus data that they drown under the weight of millions of concordance lines. Let the purely introspective tremble at a corpus-based revolution.

In 1991, with the publication of Jonathan St Clair’s *Corpus, Communication, Community*, the modern world was born. Corpus Linguists have nothing to lose but their data – but if they back up their files regularly – on external disk drives, on CD-ROMs, on USB sticks, on the Cloud – then they have a world to win.

And finally – though I hardly need to say this to those who live along the Moselle:

**Do not trust a linguist
who does not like wine.**

