

University of Surrey

Inaugural Lecture

**Linguistics: The Scientific Study
of Language**

**By Professor Corbett
8 March 1989**

LINGUISTICS: THE SCIENTIFIC STUDY OF LANGUAGE

Inaugural lecture by Professor Greville G. Corbett
8th March 1989

A chair in Linguistics and in Russian Language reflects the fact that the study of a specific language contributes to our understanding of language in general, which is the subject of Linguistics, while insights gained in Linguistics help in solving problems in analysing specific languages, such as Russian. To cover these different areas, I will first discuss the teaching of Russian at Surrey: what we do and why. Then I will consider how research is done in Linguistics, and why.¹

RUSSIAN AT SURREY

We teach Russian to students with 'A' level Russian and to beginners. Our undergraduate intake is usually between 17% and 20% of the national total for main subject Russian, and taking all the groups together there are over 140 students studying Russian at Surrey. We also put on special courses. We stress practical competence in the language and the department is known for innovation in the use of language laboratories, video, and computers. In addition students study the topics in politics, society and culture, and history which are essential for understanding the people who speak Russian.² Our students spend two periods of three months in the Soviet Union. Many go to the Lenin Pedagogical Institute in Moscow, with which the University has a beneficial exchange agreement.

WHY STUDY RUSSIAN?

Some of you will remember Professor Reeves analysing the

¹ The title was suggested to me by Professor J. R. Hurford, who gave his inaugural lecture on this topic in Edinburgh in October 1979. He too had a fine precedent since Professor Sir John Lyons' inaugural lecture at Edinburgh in January 1965 was entitled 'The Scientific Study of Language'.

² A second language may also be taken. Then, like students of French or German, students of Russian study in the new degree programme one out of Economics, European and Soviet Studies, Law and Linguistics, both as a subject in its own right and with specific reference to their particular language area.

national need for foreign language teaching,³ and some too will have heard our Vice-Chancellor explaining the benefits of foreign languages for scientists.⁴ Professor Pockney spoke on the need for Russian in 1984.⁵ The points he made then are still valid and the case for Russian is now if anything even more overwhelming.

Russian is a major language, as shown by the number of speakers (see Table 1).

Mother-tongue speakers	Official language populations
1. Chinese (1,000)	1. English (1,400)
2. English (350)	2. Chinese (1,000)
3. Spanish (250)	3. Hindi (700)
4. Hindi (200)	4. Spanish (280)
5. Arabic (150)	5. Russian (270)
Bengali (150)	6. French (220)
Russian (150)	7. Arabic (170)
8. Portuguese (135)	8. Portuguese (160)
9. Japanese (120)	9. Malay (160)
10. German (100)	10. Bengali (150)

Table 1: Speaker estimates for the world's major languages (given in millions)⁶

The first column gives the number of people for whom the language is the first language, while the second column gives the total population of countries where the language has official or semi-official status. The latter figures must be treated with particular care, since many people do not speak the official language of the country where they live; for example, the figure for English includes the 900 million inhabitants of the Indian sub-continent, even though only a small proportion of them speak

³ Professor N. Reeves, *University Language Studies and the Future of Britain*, Inaugural Lecture, University of Surrey, 1976.

⁴ Professor A. Kelly, *Science, Technology and Language* (The Threxford Lecture), published in *The Linguist*, 27, no. 2, 1988, pp. 52-58. (See pp. 55-56.)

⁵ Professor B. P. Pockney, *The Case for Russian (or Russian is a key language in any future we may have . . .)*, Inaugural Lecture, University of Surrey, 1984.

⁶ Data from D. Crystal, *The Cambridge Encyclopedia of Language* 1987, Cambridge, Cambridge University Press, p. 287.

English. In any case, the figures tell only a part of the story; Russian is the language of a superpower, on which our future depends to a large extent. Since General Secretary Gorbachev came to power, the relations of the Soviet Union with Britain and with many other countries have improved. There is now a great need for people who can perceive the problems and opportunities of the next century; understanding and working with the Soviet Union will be one of them. Furthermore, the reduced predictability which Gorbachev brings also involves certain greater risks, and having an adequate number of people competent in the language and with an understanding of current Soviet affairs is essential. A lesson we could learn from the Falklands is that training sufficient people in languages may save lives and money.

The need for Russian is recognised in some quarters. Students recognise it, since applications for Russian rose last year.⁷ And the Foreign Affairs Committee of the House of Commons gave the following judgement:⁸

' . . . the Helsinki Final Act committed signatory states not only to encourage the study of foreign languages and civilisations but also specifically to "stimulate . . . the diversification of choice of languages taught at various levels, paying due attention to less widely-spread or studied languages." In recent years the British Government appears to have sadly ignored these principles and commitments, at least so far as the Soviet Union is concerned. **At a time of increased awareness of the need to improve relations with the countries of Eastern Europe, we believe that HM Government should now seek to honour these commitments with a view to establishing the study of the Russian language and of the Soviet Union as a greater priority at all levels of the educational system.**'

Even the Minister deplores the decline in provision for Russian, while ignoring the fact that this decline has to be set against a

⁷ E. Williams in the *Times Higher Educational Supplement*, 12 August 1988, p. 1.

⁸ Foreign Affairs Committee 1985-86, *UK-Soviet Relations: Volume 1: Report, Appendices and Minutes of Proceedings: Second Report of the Session*, 1986, London, Her Majesty's Stationery Office, p. lxiii. (Bold print in original.)

substantial reduction in Government funding for the universities. We must be thankful that Russian has been able to continue here, and has been encouraged to flourish in some areas. But even at Surrey, as a result of continuing government cuts, we cannot carry out our teaching and research at the level we would wish, and this in areas which the Commons Committee believes deserve greater priority. Of course, this frustration is not unique: it is shared by many colleagues in all disciplines throughout the University sector.⁹

I believe that in studying the Soviet Union and the Russian language, in producing students with such a background, in arranging to send large groups to the Soviet Union and in receiving groups here, people at Surrey are making the 21st century slightly more likely to occur.

RESEARCH IN LINGUISTICS

Despite these excellent reasons for an interest in Russian, my prime motivation is rather different; it is the attempt to understand language. To that end, Russian provides fascinating subject matter. In its linguistic, rather than its socio-political context, Russian is one of the Slavonic languages. To gain an impression of this group, consider the following sample of each.¹⁰

⁹ As Sir George Porter put it: '[knowledge] is a spiritual power and a part of the human purpose. A country that lets it languish becomes impoverished in more than its Treasury.' (In *Knowledge Itself is Power*, The Richard Dimbleby Lecture 1988, published by BBC Books, London, 1988, p. 28.)

¹⁰ John 3.16, from *The Gospel in Many Tongues: Specimens of 875 Languages*, 1965, London, British and Foreign Bible Society.

RUSSIAN

Ибо так возлюбил Бог мир, что дал Сына Единородного, чтобы каждый верующий в Него не погиб, но имел жизнь вечную.

BELORUSSIAN

Бо так пасябѣў Бог свет, што даў Сина Свѣдо-
Адынароднага, каб усеякі, хто веруе ў Яго, не згінуў,
але меў жыццѣ вѣчныя.

UKRAINIAN

Так бо Бог полюбив світ, що дав Сина Свого Одно-
родженного, щоб кожен, хто вірує в Нього, не загинув,
але мав життя вічне.

Table 2: Samples of the East Slavonic languages

Note the use of the Cyrillic script for these languages. This should be contrasted with the script of the West Slavonic languages:

POLISH

Albowiem tak Bóg umiłowal świat, że Syna swego
jednorodzanego dał, aby każdy, kto weń wierzy, nie
zginął, ale miał żywot wieczny.

CZECH

Nebo tak Bůh miloval svět, že Syna svého jedno-
rozeného dal, aby každý, kdož věří v něho, nezahynul,
ale měl život věčný.

SLOVAK

Lebo tak miloval Bůh svet, že svojho jednorodného
Syna dal, aby každý, kto verí v neho, nezahynul,
ale mal večný život.

LOWER SORBIAN

Wšakto tak jo Bog ten swjet lubowal, aŝ won swojogo
jednoroźnjaknego syna dal jo, aby šakřitne do noga
wřetnje, šubjenje nebuŝi, ale to ništernje ŝiwjenje mjeŝi.

UPPER SORBIAN

Wšakto tak je Boh ten swjet lubowal, ŝo won
swojeho jeničjeho naroźnjaknego Syna dal je, ŝo bogaŝe
šakřit, ŝiŝ do njego wjerje, šubeni nebuŝi, ale mjeŝtne
ŝiwjenje mjeŝi.

Table 3: Samples of the West Slavonic languages

Here we find the Roman script. The examples from Upper and Lower Sorbian are in the Gothic typeface. These two texts are from 1860; a contemporary text would have the more familiar Roman typeface. Finally we should consider the South Slavonic languages:

BULGARIAN

Зашто Богъ толкозъ възлюби свѣтътъ, што даде
свој единороденъ Синъ, за да не погине ни слншъ, който
вѣрва въ него, но да има вѣченъ животъ.

MACEDONIAN

Зашто Бог толку го возљуби светот, што Го даде
Својот Единороден Син, та секој што верува во Него
да не погине, но да има живот вечен.

SLOVENIAN

Kajti tako je Bog ljubil svet, da je dal Sina svojega
edinorojenega, da se ne pogubi, kdorkoli veruje vanj,
temuč da ima večno življenje.

SERBO-CROAT: Eastern variant

Јер је Бог тако заволео
свет, да је и Сина свога јединорођенога дао,
да виједан који верује у њега, не погине, него
да има живот вечни.

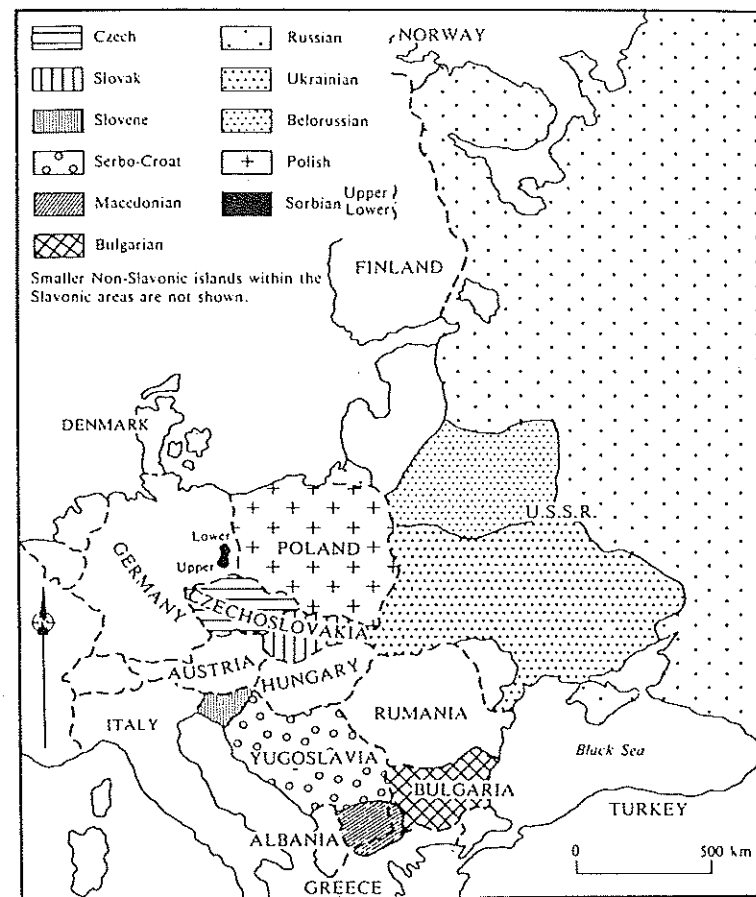
SERBO-CROAT: Western variant

Јер Бог толико љуби свет, да је и Сина свога јединоро-
ђенога дао, да ни један који у њега верује не пропадне,
него да има живот вечни.

Table 4: Samples of the South Slavonic languages

Both Cyrillic and Roman scripts are shown here. The distribution of these scripts has a historical explanation, as can be seen from the areas where the different languages are spoken (see Map 1). The languages spoken to the east, in orthodox countries, use the Cyrillic script; that is, Russian, Belorussian, Ukrainian, Bulgarian, Macedonian, and the Eastern Variant of Serbo-Croat. The languages spoken to the west, in catholic countries (later protestant in some cases), use the Roman script; these are Polish, Czech, Slovak, Upper and Lower Sorbian, Slovenian, and the

Western Variant of Serbo-Croat. We can see that different languages may share a script (like Czech, Polish and English) and that one language may use two different scripts (Serbo-Croat).



Map 1: Distribution of the Slavonic languages¹¹

¹¹ Map from B. Comrie (editor) *The World's Major Languages*, 1987, London, Croom Helm, p. 323, courtesy of Routledge publishers.

Thus the subject matter of Linguistics is vast. Yet some cannot see how research on language can be done. The problem is that of familiarity. We all use language; almost everyone acquires a language (only children with severe problems do not). And in many parts of the world it is normal to learn several languages. So what is there to investigate? Here an analogy may be helpful. Considerable visual skills are required to steer a car through traffic; but those who have driven here this evening would not claim to understand how vision works. Being able to do something is not the same as understanding how it is done.

Imagine we have a Martian visitor; it has good processing skills and has learned elements of English. It cannot yet, however, ask 'yes-no' questions. Could we give instructions? If we have:

(1) Mary can speak Russian.

We can tell our Martian to form a yes-no question by putting the second word first (though this assumes that it knows what a word is):

(2) Can Mary speak Russian?

This seems to work:

(3) John is at home reading Dostoevsky.

(4) Is John at home reading Dostoevsky?

Unfortunately our Martian will have problems with such a rule:

(5) Your friend can speak Russian.

(6) *Friend your can speak Russian?

(An asterisk indicates an unacceptable sentence.) We should try a more complex solution - the verb is put at the front of the sentence; this requires that we explain what verbs are, or list them. Then we would obtain (7) instead of (6):

(7) Can your friend speak Russian?

With this rule our Martian could handle all the examples given so far, but not:

(8) Your friend speaks Russian.

According to the rule suggested, the question ought to be:

(9) *Speaks your friend Russian?

So we must say that certain verbs (auxiliaries) occur in initial position; these include *do*, which is placed there for forming questions when no other auxiliary is present:

(10) Does your friend speak Russian?

Furthermore we need to specify which auxiliary verb occurs in initial position:

(11) Your friend, who is studying French, can also speak Russian.

(12) *Is your friend, who studying French, can also speak Russian?

Such a question would certainly give away our Martian as not being a native speaker of English. It might be argued that no-one in their right mind would produce a question like (12); indeed not - this illustrates well just how much we 'know' about language, but at a subconscious level. Even if we could not externalize the information, we know that the auxiliary in initial position must be that of the main clause:

(13) Can your friend, who is studying French, also speak Russian?

This area of English syntax turns out to be far from simple.¹⁴ While we may tutor Martians only rarely, the need for an unambiguous description of language is real: we wish to program computers to use language. They can carry out processing rapidly, but they do not have the human's knowledge of natural language; they do not behave as though they had a 'right mind'.

We have seen that language has regularities and structures which are not immediately obvious. How then do we investigate it? I

¹⁴ For a helpful discussion see N. Smith & D. Wilson, *Modern Linguistics: The Results of Chomsky's Revolution*, 1979, Harmondsworth, Penguin, pp. 85-9, and for a formal account see G. Gazdar, E. Klein, G. K. Pullum and I. A. Sag *Generalized Phrase Structure Grammar*, 1985, Oxford, Blackwell, pp. 60-65.

suggest that research in Linguistics is carried out much as in any other science. The first step is to identify a problem, to recognise that something is strange. The story, better called the parable, about Newton is instructive. Newton sees an apple fall and realises that this is a problem, something requiring a proper explanation.¹⁵ There is the insight that things might be different, that objects might remain suspended in the air, bounce perpetually and so on.

"Discovery consists of seeing what everybody has seen and thinking what nobody has thought." (Albert Szent-Gyorgyi)

In Linguistics the very familiarity of language tends to hide its strangeness. As Chomsky puts it: 'Phenomena can be so familiar that we really do not see them at all . . .'.¹⁶ Looking at a foreign language helps to give distance. Suppose we ask what the Russian words for 'one', 'two', 'three', 'four' and 'five' are. The answer is surprising:

one	two	three	four	five
один	два	три	четыре	пять
одна	две	трѐх	четырёх	пяти
одно	двух	трѐм	четырёх	пятью
одну	двум	тремя	четырьмя	
одного	двумя			
одной				
одному				
одним				
одном				
одни				
одних				
одним				
одними				

Table 5: The numerals 1-5 in Russian

¹⁵ Of course, Newton was not the first to offer a theory as to why objects fall to earth.

¹⁶ N. Chomsky *Language and Mind*, enlarged edition, 1972, New York, Harcourt Brace Jovanovich, p. 24.

It may be easier if we look at the same forms in transliteration:

one	two	three	four	five
odin	dva	tri	četyre	pjat'
odna	dve	trëx	četyrëx	pjati
odno	dvux	trëm	četyrëm	pjat'ju
odnu	dvum	tremja	četyr'mja	
odnogo	dvumja			
odnoj				
odnomu				
odnim				
odnom				
odni				
odnix				
odnim				
odnimi				

Table 6: The numerals 1-5 in Russian (transliterated)

The first surprise is that there are so many forms. The second surprise is that different numerals have different numbers of forms. Let us investigate how some of them are used. If we take a single construction and use the numerals with different nouns we obtain the following data:

- (14) *odin žurnal* *odna gazeta* *odno okno*
 one magazine one newspaper one window

We see that the form of *odin* 'one' differs according to the noun. In fact the nouns of Russian can be divided into three groups or genders and *odin* (and other such words) alters its form according to the noun; it is said to 'agree' with it in gender. When we look at *dva* 'two' the picture changes:

- (15) *dva žurnala* *dve gazety* *dva okna*
 two magazines two newspapers two windows

(Literally it is 'two of a magazine'.) With *dva* we can account for

only two different forms instead of three in terms of agreement in gender. Now look at *pjat'* 'five':

- (16) *pjat'* žurnalov *pjat'* gazet *pjat'* okon
 five magazines five newspapers five windows

Here *pjat'* does not change; but the forms of the nouns are different again. A literal translation would be 'five of magazines'. So we cannot account for the forms of *pjat'* in terms of gender, but we have a new problem to solve (the different forms of the nouns). Each factor which we find determines some of the forms must be carefully tested on each numeral and construction. Then we must organise the data to give an explicit account of the phenomenon. This is done in Table 7.

	odin	dva	tri	<i>pjat'</i>	sto	tysjača	million
	1	2	3	5	100	1,000	1,000,000
1. agrees with noun in syntactic number	+	-	-	-	-	-	-
2. agrees in case in the direct cases	+	-	-	-	-	-	-
3. agrees in gender	+	(+)	-	-	-	-	-
4. agrees in animacy	+	+	+	-	-	-	-
5. has no semantically independent plural	+	+	+	+	(-)	-	-
6. fails to take agreeing determiners	+	+	+	+	+	-	-
7. does not take noun in genitive plural throughout paradigm	+	+	+	+	+	±	-

Table 7: The simple cardinal numerals of Russian

Along the top we have the simple cardinal numerals of Russian; on

the left we have various tests we may apply to these numerals, the third of which was illustrated in the examples we examined. The tests are devised such that we would expect an adjective, a word like *krasivyj* 'beautiful', to pass them, while a noun, a word like *stado* 'flock', should fail. So we now have a more explicit statement of the problem. By ordering the numerals and the tests as in Table 7, we obtain a highly regular picture. We see that each numeral passes successively fewer tests - each is less like an adjective and more like a noun than the one to its left. Regular matrices of this kind will be recognised by scientists from various disciplines. They have the property that the values decrease monotonically (i.e. without intervening increase) across the rows and up the columns. Their importance in Linguistics was recognised by Ross.¹⁷

We can say then that the behaviour of Russian numerals is not random. The numerals fall between the poles of adjective and noun. But there is another vital point about the matrix. The numerals were arranged so as to give a regular pattern in linguistic terms. And that turns out to be the same as numerical order. But *a priori* there was no reason to expect that the numerical value of a numeral would be important for linguistic analysis. Our hypothesis then, which covers the data in question, is that the syntactic behaviour of Russian numerals falls between that of adjectives and nouns, and that the higher the numeral is, the more like a noun it will be.

Having found such a remarkable and unexpected regularity we might predict that other languages will behave similarly. And indeed we find the same pattern constantly repeated. Not in the luxuriant detail of Russian, but repeated nevertheless. For example, in the Bantu language Chichewa (spoken in Malawi), we find:

- (11) *chi-pewa* *chi-modzi*
 hat one 'one hat'

Here the numeral *-modzi* 'one' shows agreement in gender with the noun. The numeral *-wiri* also agrees in gender (it agrees in number too and is plural, like the noun):

¹⁷ J. R. Ross 'The Category Squish: Endstation Hauptwort' *Papers from the Eighth Regional Meeting Chicago Linguistic Society*, 1972, pp. 316-328, and subsequent papers.

(12) zi-pewa zi-wiri
 hats two 'two hats'

In fact all the numerals from one to five agree like adjectives; but *khumi* 'ten' does not agree - its behaviour is more like that of a noun:

(13) zi-pewa khumi
 hats ten 'ten hats'

This is further demonstrated by the fact that it has a plural *makhumi* 'tens'. And in English, we have *one, two, three . . . books* but *a hundred, a thousand, a million books*. It appears to be universally true that if simple cardinal numerals vary in their syntactic behaviour it is the higher which will be more noun-like.¹⁸

To recap then: we began with a phenomenon we recognised to be a problem. Next we assembled relevant data, describing it explicitly, and we framed a hypothesis to cover the data. Checking the hypothesis leads to further data collection and reformulation of the hypothesis. This process is lengthy and untidy. Despite attempting to break Bloomfield's injunction ('Don't take your guests into the kitchen')¹⁹ I still cannot convey just how chaotic that stage can be. Having reached a hypothesis which covered the facts of Russian, we were able to derive predictions from it and to check them. The procedure is represented in Figure 1.

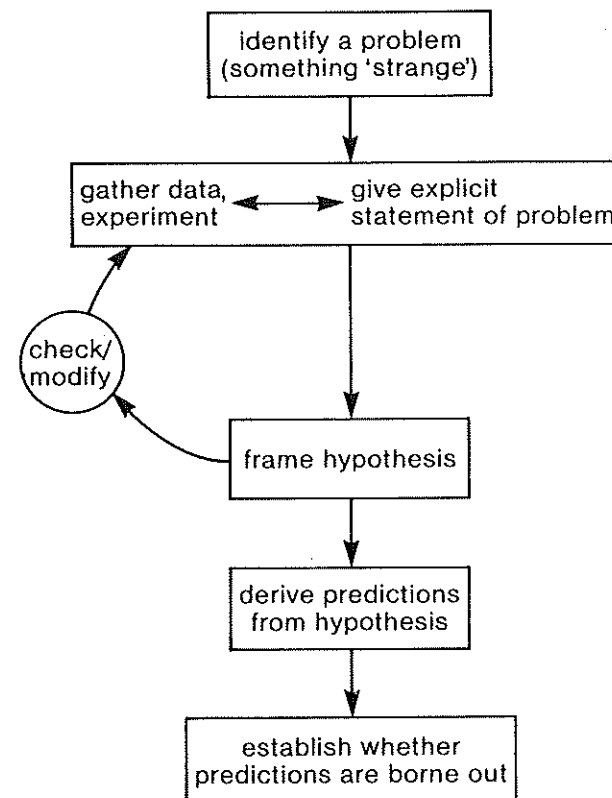


Figure 1: Linguistic research (first approximation)

This seems straightforward - it is misleadingly so. Let us examine why. At the first stage we found something strange. Relevant to this is one of anon's little gems:

'Life is very strange' said Jeremy.
 'Compared with what?' replied the spider.²⁰

The point is that to say that a phenomenon is strange we must have a point of comparison, a theory in fact, even if we hesitate to use the term. Our theory might have been that all languages are like English (or should be). Or that numerals are relatively uncomplicated linguistically; or that words of a single type (like

¹⁸ G. G. Corbett 'Universals in the syntax of cardinal numerals' *Lingua* 46 (1978), p. 368. The original formulation has been tightened up by J. R. Hurford *Language and Number: The Emergence of a Cognitive System*, 1987, Oxford, Blackwell, pp. 187-197.

¹⁹ Quoted by C. F. Hockett 'What Algonquian is really like' *International Journal of American Linguistics* 32 (1966), p. 67.

²⁰ Quoted in A. L. Mackay *The Harvest of a Quiet Eye: A Selection of Scientific Quotations*, 1977, Bristol and London, The Institute of Physics, p. 2.

numerals) will behave in a similar fashion. However rudimentary it may have been, theory was involved at the start (see Figure 2).

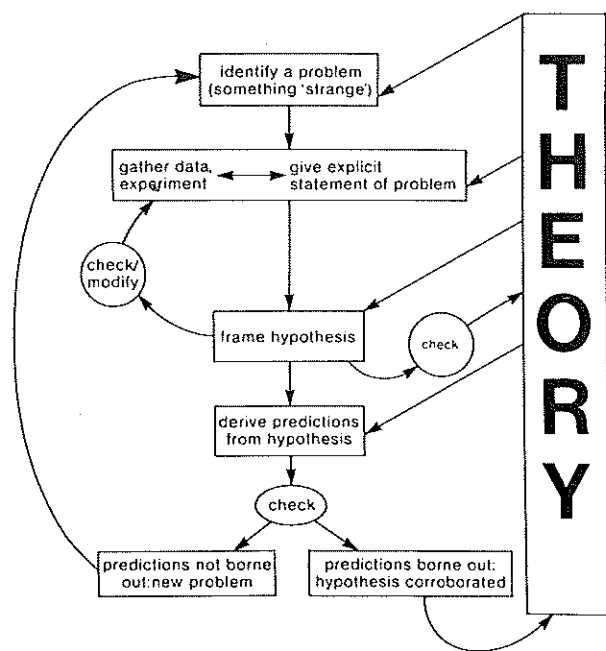


Figure 2: Linguistic research (recognizing the rôle of theory)

In stating the problem clearly, we used terms like noun and adjective. Though basic, these are theoretical constructs. In framing our hypothesis there was an even clearer theoretical input, since the type of regular matrix arrived at had been proposed in the linguistic literature (though such a regular example had not to my knowledge been found exemplified in natural language). There is a second interaction with theory here, since a hypothesis is checked not only against the data, but also for consistency both internally and with other parts of linguistic theory.²¹

The predictions we draw from our hypothesis also depend on our

²¹ Compare K. A. Popper *The Logic of Scientific Discovery*, 1959, London, Hutchinson, pp. 32-33.

theoretical assumptions. The jump from an analysis of one language to claims about other languages depends on the notion that languages are in some sense the same, a claim made by Roger Bacon in the thirteenth century.²² If our hypothesis is corroborated this too has implications for theory. In our case it added a claim about numerals in language (thus contributing a small part to our definition of 'possible human language'); and in more general terms, it is consonant with work on prototypes, notably by Rosch, which has had a considerable impact in Psychology and Linguistics. And if the predictions are **not** borne out, then we have something strange, and the process begins again. Thus, though it may not have been immediately obvious, theory was involved at every stage.

While analysing numerals we met the category of gender. This category appears strange to speakers of languages where gender plays a minimal rôle or is absent, and so it deserves a further look.²³ We saw that Russian nouns can be divided into three groups; here are some more examples:

otec	father	mat'	mother	okno	window
djadja	uncle	tetja	aunt	moloko	milk
byk	bull	korova	cow	vremja	time
žurnal	magazine	gazeta	newspaper	pis'mo	letter
avtomobil'	car	mašina	car	taksi	taxi
flag	flag	èmblema	emblem	znamja	banner
zakon	law	glasnost'	openness	doverie	trust

masculine feminine neuter

Table 8: Examples of nouns of the three genders in Russian

²² "Grammatica una et eadem est secundum substantiam in omnibus linguis, licet accidentaliter varietur." (Grammar is one and the same in all languages in substance, though it may vary in accidents.) Quoted by C. A. Ferguson 'Historical background of universals research' in J. H. Greenberg, C. A. Ferguson and E. A. Moravcsik (editors) *Universals of Human Language: I: Method and Theory*, 1978, Stanford, Stanford University Press, p. 9. Those who would ally themselves to Bacon's claim today have to take a more abstract view of language than he is likely to have intended.

²³ The sources for the data in this section are given in G. G. Corbett *Gender to be published by Cambridge University Press.*

How does a native speaker know which nouns belong in which group? It turns out that the information about each noun which has to be remembered in any case is sufficient to allow the gender to be derived. In some instances meaning determines gender: *otec* 'father', *djadja* 'uncle', *byk* 'bull', are masculine; *mat'* 'mother', *tetja* 'aunt', and *korova*, 'cow' are feminine. These 'sex-differentiable' nouns are those where the sex of the referent matters to humans (the sex of other humans, and of animals particularly for breeding and milking). But this rule will not account for the gender of the remaining nouns in our list. Here there is information about the different forms of the words, their morphology, which allows their gender to be determined. Let us look at the algorithm which accounts for gender assignment in Russian (Figure 3).

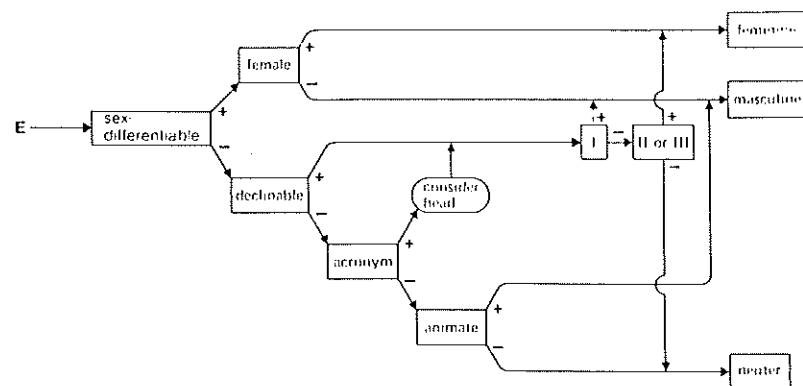


Figure 3: Gender assignment in Russian

The analysis requires that I postulate four rather than three declensional classes (indicated by Roman numerals, the fourth being covered by the negative exit from the decision box which includes II and III). Russianists will realize that this represents a radical break with tradition. But the details need not concern us here. What matters is that two types of information are required: the meaning (semantics) and the various word-forms (morphology) of the noun. When the two conflict (as in the case of *djadja*, 'uncle') then meaning takes precedence.

When we try to generalize our result, we soon find that not all systems are identical; however, the factors found in Russian give us a possible typology. Meaning alone is sufficient to determine gender in some languages. For example, in Tamil (a Dravidian language of southern India and Sri Lanka), nouns denoting gods and male humans are masculine, those denoting goddesses and female humans are feminine, and all others are neuter. Even without knowing the form of the noun one can be confident that the word for a man is masculine, for a woman is feminine while that for an elephant is neuter. Some other Dravidian languages like Kolami have only two genders: male humans are masculine and all other nouns fall into the non-masculine gender. Sometimes this situation is reversed. In Diyari, a language of South Australia, there are again two genders: one is for nouns with female referents (women, girls, doe kangaroos), and the other is for all remaining nouns. And in Dizi, an Omotic language of south-west Ethiopia, there is one gender for nouns denoting females (humans and animals) and for diminutives; all remaining nouns are masculine. In Alambak, a Sepik Hill language of Papua New Guinea, the masculine gender includes nouns denoting males and those denoting things like crocodiles, pythons and arrows, which are long and thin, while the feminine is for nouns denoting females or short, squat items like turtles, frogs and chairs.

The semantic criteria by which nouns are assigned to genders may be less straightforward. Thus Dyrbal, a language of North Queensland, Australia, has four genders, primarily for:

1. male humans and non-human animates
2. female humans
3. non-flesh food
4. others

There are many apparent exceptions. For example the moon is in the first, masculine gender and the sun is in the second, feminine gender. The reason is that in Dyrbal mythology, as indeed in much of Australia, the moon is the husband of the sun; in Dyrbal the rôle in mythology determines gender. World view also plays a part in Ojibwa. Here, as in other Algonquian languages, there are two genders: animate and inanimate. The first includes nouns denoting persons, animals, spirits and trees. But it also contains some surprises, such as the nouns for 'sacred story', 'star', 'pipe' (for smoking) and so on. It has been suggested that the animate nouns are in fact nouns denoting objects which in the world-view of the Ojibwa are sources or carriers of 'power'.

These then were all languages in which the meaning of the noun determines gender. We noted a second possibility earlier: in

Russian meaning is not sufficient - it has to be supplemented by morphological information. The same is true of many Indo-European languages, and a similar system is found in Bantu languages. Note that this typological approach cuts across the genetic classification of languages.

The third possibility is that gender is determined by meaning or, when that fails, by the sound-shape or phonology of the noun. We find this system in various languages around the world; a particularly interesting example is Qafar, a Cushitic language of north-east Ethiopia and Djibouti. The semantic rule in Qafar is straightforward: nouns denoting males are masculine; those denoting females are feminine. Thus *bāq̄la* 'husband' is masculine while *barrà* 'wife' is feminine. For nouns which do not denote sex-differentiables there is a simple but very interesting rule. Nouns ending in a vowel which can bear high tone are feminine: for example, *karmà* 'autumn'. All others are masculine: *gilàl* 'winter' ends in a consonant and so is masculine, while *baànta* 'trumpet' ends in a vowel but not one bearing high tone, and so is also masculine.

We have thus identified three types of assignment system:

- I meaning only
- IIa meaning and form (morphology)
- IIb meaning and form (phonology)

Table 9: Typology of gender assignment systems

This is a good opportunity to see how linguistic statements can go beyond description to prediction. I claim that gender assignment is essentially systematic in all languages. This is a radical claim since the prevailing view has been that it is not. And yet, to date, it appears justified. Second, on the basis of our typology we predict that no language has a gender system based on form only; this too is a claim readily open to falsification, but to date it holds true. And at a lower level we can make predictions for specific languages. Languages frequently borrow new words from other languages. This serves as a continuously running experiment. We predict that borrowed nouns will take their gender according to the rules given. For example, in Tamil, the borrowed word *ḍaekṭar* 'doctor' is masculine or feminine, depending on the sex of

the referent, while *kaaru* 'car' is neuter.

We can go further by constructing words which do not actually exist. Their gender is predicted by our rules and so we can test our predictions. I have been fortunate to work with Dr. Hayward on Qafar: he constructed nouns to test the rules given for Qafar; and indeed a native speaker assigned each to the gender predicted (but had no idea how he did it).

I mentioned earlier that while collecting data we also have to restate the problem more clearly. For the kind of work described it is important to be sure that we are comparing like with like. In the case of gender systems, definition has often been a problem. To define gender we start from the notion of agreement class:

An agreement class is a set of nouns such that any two members of that set have the property that whenever:

- (i) they stand in the same morphosyntactic form, and
- (ii) they occur in the same agreement domain, and
- (iii) they have the same lexical item as agreement target,

then their targets have the same morphological realization.

This notion must be further restricted, and an account of the problems involved would take us too far afield. Let us simply retain the point that to demonstrate the existence of a gender system we need agreement evidence, that is, evidence outside the noun itself, and let us look at a concrete example. Archi is a language spoken by under 1000 people in the Dagestan highlands. (The Caucasus is known as the mountain of languages; there are almost 40 languages which share only 6 million speakers.) Archi is claimed to have four genders (for male humans, female humans, a third gender for most animals and some inanimates, and a fourth for the residue, that is, young and some smaller animals, and the remaining inanimates). The type of evidence available is given in the following four sentences:

- (14) w-ez dija k'anši w-i
 I father like is 'I like father'
- (15) d-ez buwa k'anši d-i
 I mother like is 'I like mother'
- (16) b-ez dogi k'anši b-i
 I donkey like is 'I like the donkey'
- (17) Ø-ez motol k'anši Ø-i
 I young goat like is 'I like the kid'

This is an ergative language: the part of the verb which shows agreement agrees with what we should think of as the object. With verbs of emotion and perception, the subject stands in the dative case; it is a personal pronoun with an agreement slot, and so also agrees with the object. Thus there is agreement evidence for four genders. If however the nouns are plural, there are only two distinct agreement forms, one for genders one and two and one for genders three and four. It seems therefore that there may be interesting patterns in the genders as they occur on agreement targets such as verbs. If we take the (prefixal) agreement markers of Archi, we find the following pattern:

	singular	plural	singular	plural
I (father)	w	b	w	b
II (mother)	d	b	d	b
III (donkey)	b	φ	b	φ
IV (kid)	φ	φ	φ	φ

Figure 4: The gender pattern of Archi

On the left we have just the agreement markers. These are given again on the right in a notation which recognises the identities of form. The lines joining the agreement markers represent classes of nouns, and these are labelled with Roman numerals. At first sight this pattern is surprising; in part this may be because those who have learned a gender language have probably learned French, a

language which has a simpler pattern. In French, if a noun is masculine, it takes masculine agreements in singular and plural. The pattern is as follows:

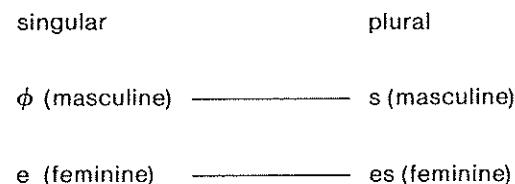


Figure 5: The gender pattern of French

We call systems like this 'parallel':

A parallel system is one in which gender in one number determines gender in the other and vice versa (there is a one-to-one mapping of the target genders in one number on to the target genders in the other).

The next type, as illustrated by Archi and, nearer to home, by German, is called 'convergent':

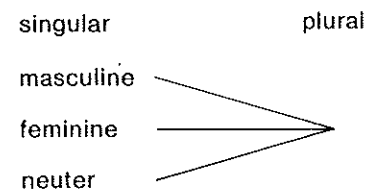


Figure 6: The gender pattern of German

We define a convergent system as follows:

A convergent system is one in which gender in one number determines gender in the other but not vice versa (there is a many-to-one mapping of the target genders in one number on to the target genders in the other).

The most complex type is the 'crossed' type:

A crossed system is one in which gender in neither number determines gender in the other (there is what some call a 'many-to-many' mapping between the target genders in the different numbers).

A good illustration is provided by Rumanian:

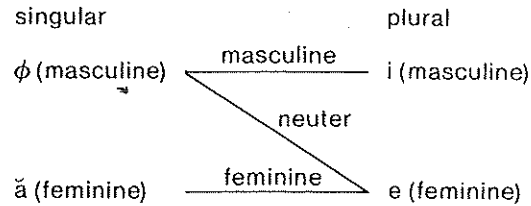


Figure 7: The gender pattern of Rumanian

The systems may be complicated by the presence of subgenres, as in the case of Serbo-Croat:

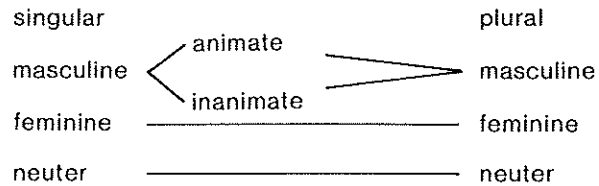


Figure 8: The gender pattern of Serbo-Croat

Russian too has subgenres, but they are more fully integrated into the gender system:

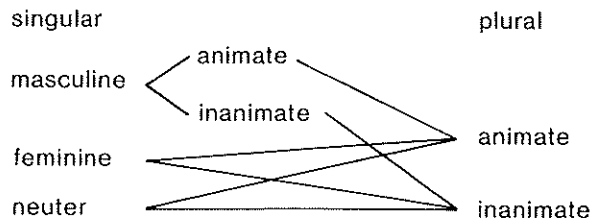


Figure 9: The gender pattern of Russian

Upper Sorbian presents a subgenre in a three number system:

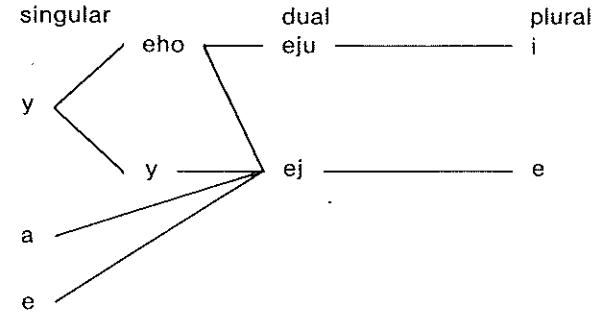


Figure 10: The gender pattern of Upper Sorbian

And the most complex crossed system found to date is that of Fula (a member of the West Atlantic branch of Niger-Kordofanian):

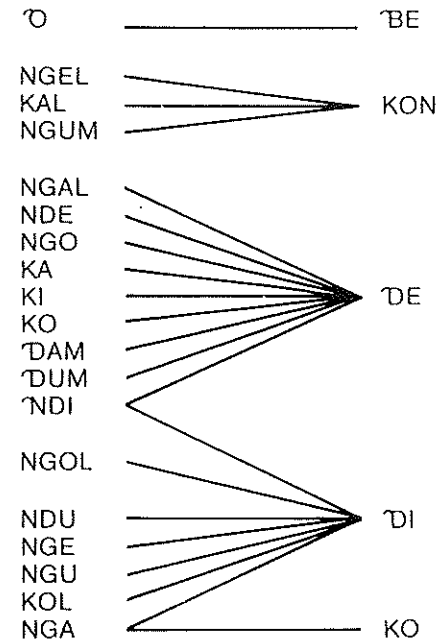


Figure 11: The gender pattern of Fula

The schemes allow us to see clearly how gender systems grow and

decline. A striking example of how they grow is provided by the various dialects of Andi (North-East Caucasian). The most conservative dialects, which we shall call 'Type A', preserve a very interesting system, which is of the parallel type:

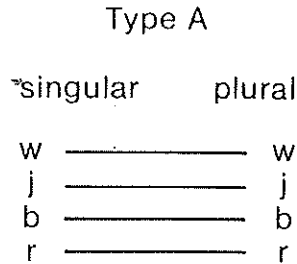


Figure 12A: The gender pattern of Andi: dialect type A

When we turn to Type B, the most common type of Upper Andi dialect, we note an apparently minor change; some nouns which earlier took *b* as their sole agreement marker now take *b* only when singular but take *j* when plural.

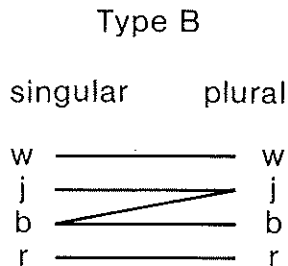


Figure 12B: The gender pattern of Andi: dialect type B

This change means that type B dialects have a crossed rather than a parallel system and that they have five genders rather than four. It is interesting to look at the gender assignments involved. The first two genders (taking *w - w* and *j - j*) are for male humans and female humans respectively. The original third gender

(*b - b*) included most of the animates together with some inanimates. In type B this gender has been split, the animates forming the new gender (*b - j*) and the inanimates remaining as the (*b - b*) gender. Type C, represented only by the Rikvani dialect, has gone a stage further:

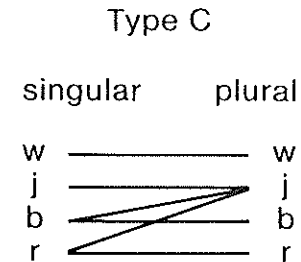


Figure 12C: The gender pattern of Andi: dialect type C

The smaller number of animates which were in the original fourth gender (*r - r*), mainly denoting insects, were separated from the other nouns, which denote inanimates, to form a sixth gender (taking the agreement markers *r - j*). The system is again crossed, being somewhat more complex than that of the type B dialects.

The way in which genders can be lost is illustrated by various Dravidian languages. The earliest establishable system appears to be that still found in languages like Tamil. It is a convergent system, as shown in Figure 13A:

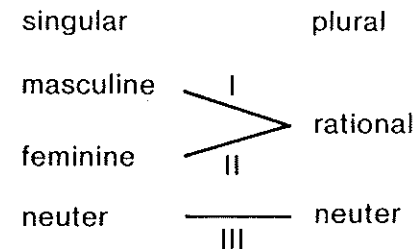


Figure 13A: Gender in Tamil

Nouns denoting male rationals (I) take masculine forms in the singular, and rational in the plural; nouns denoting female rationals (II) take feminine and rational forms, while nouns denoting non-rationals (III) take neuter agreement in both numbers. Languages like Telugu show a system one stage on from this (Figure 13B):

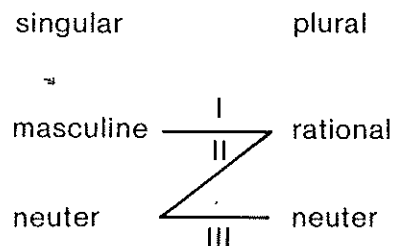


Figure 13B: Gender in Telugu

As compared with Tamil, Telugu divides its nouns into the same three genders (I - male rational, II - female rational, and III - neuter), but it has one less gender agreement form. The loss of this form has changed a convergent system into a crossed one. The last stage is found in languages like Kolami (Figure 13C)

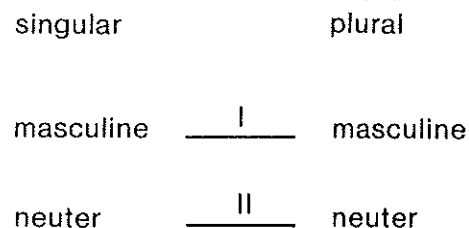


Figure 13C : Gender in Kolami

In the Kolami development, all nouns taking neuter agreement in the singular take neuter agreement in the plural. This means that we now have a parallel system, and one with only two genders: I - male personal, and II: all others (as mentioned earlier). A pattern which remains constant in all these languages is that there are never more gender distinctions in the plural than in the singular, an observation due to Greenberg.²⁴

²⁴ J. H. Greenberg 'Some universals of grammar with particular reference to the order of meaningful elements' in J. H. Greenberg (editor) *Universals of Language*,

So far we have taken the data for granted, but the collection of data can be time consuming and difficult. There are two main methods: informant work and studies based on a corpus (a collection of spoken or written material). I will give an illustration of each. A project where informant work played a major rôle began after I learned about the following construction in Upper Sorbian (a Slavonic language spoken in the eastern part of Germany). It ties in with the previous analysis of gender:

- (18) *moj-eho* *bratr-ow-e* *džěć-i*
 my-GEN.SG.MASC brother-POSS-NOM.PL children-NOM.PL
 'my brother's children'

This example also illustrates the rôle of theory well. Linguists react to (18) with surprise and disbelief. But to those who have no views about syntax and morphology for which they would go to the stake, (18) may not seem strange. Let us first examine the different elements of the construction. *Bratrowe* is a possessive adjective formed from the noun *bratr* 'brother'. The adjectival stem *bratrow-* takes the ending *-e* to show agreement with the head noun *džěći*. The problem is *mojeho* which has no apparent head; clearly it does not agree with *džěći*, since it carries the wrong features. It seems rather that its agreement controller is a form of *bratr*, which is masculine singular and underlies the possessive adjective.

Why then is the existence of the construction illustrated in (18) so strange? Put simply, we expect phrases to be analysable into smaller phrases and finally into words in a hierarchical fashion, and for some words to be further analysable. In this example, the phrase structure and word structure do not match, as can be seen from (19), where the internal phrases are indicated by brackets:

- (19) [*moj-eho* [*bratr*]-ow-e *džěć-i*]

A second reason why it is strange is that the construction is not found in some other closely related Slavonic languages. The first step then is to check. I went to talk to native speakers, and confirmed that they accept examples of this construction. Two related constructions also occur:

1966 (second edition), Cambridge, Massachusetts, MIT Press, p. 112.

(19) Słyšetaj...Wićazowy hłós, kotryž je zastupił
 (They) hear Wićaz's voice, who is gone in
 'They hear the voice of Wićaz, who has gone in'.

The relative pronoun *kotryž* is masculine singular; the sense shows that its antecedent is *Wićaz*, the noun underlying the possessive adjective, and not *hłós* (which is also masculine singular).

The possessive adjective can similarly control personal pronouns:

(20) To je naš-eho wučerj-ow-a
 that is our-GEN.SG.MASC teacher-POSS-NOM.SG.FEM
 zahrodka. Wón wjele
 garden.NOM.SG.FEM he.NOM.SG.MASC a lot
 w njej dźěła.
 in it works
 'That is our teacher's garden. He works in it a lot.'

The personal pronoun *wón* takes as its antecedent the noun phrase *naš wučer* 'our teacher', which underlies the phrase headed by the possessive adjective. The next step was to see whether these constructions occurred in all the languages which could conceivably have them, that is all the Slavonic family. To find out whether a construction is impossible, traditional sources are of limited help. So over a period of two years I gathered the judgements of some fifty native speakers. The results are given in Table 10.

	attributive	relative pronoun	personal pronoun
East Slavonic			
Russian	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Russian (19th century)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Old Russian	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Belorussian	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Ukrainian	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
South Slavonic			
Bulgarian	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Macedonian	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Slovenian	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Serbo-Croat	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Old Church Slavonic	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
West Slavonic			
Polish	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Czech	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Slovak	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Lower Sorbian	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Upper Sorbian	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Table 10: Control Possibilities of the Possessive Adjective in the Slavonic Languages

This table gives data on all the modern Slavonic languages, with some historical data too. The generalizations are clear. Informants accept control of the attributive modifier (as in the first example) only if the second construction (control of the relative pronoun) is accepted, and that only if control of the personal pronoun is possible. The hierarchy of three positions just given can be subsumed under an already postulated hierarchy (the Agreement Hierarchy) so the new claims amount to relatively little. Since the Slavonic languages are closely related, they act like experimental controls on each other. A good deal is held constant, while the points of special interest here show variation.

The second major source of data is linguistic corpora. Using this approach allows the linguist to uncover subtle effects which are difficult to demonstrate using informant work. For illustration, I will present a construction in Russian (predicate agreement with conjoined noun phrases) which allows two alternative forms. I believed there were two independent factors involved (the animacy of the head nouns and the position of the subject, before or after the predicate). There are thus four possibilities. The details need not concern us - let us simply call the possibilities A, B, C, D. I then collected a corpus of Russian texts, from the beginning of the last century up to the late 1970's, of some one million words. I read them, recording every example which they contained (726 in all). The results are presented graphically in Figure 14.

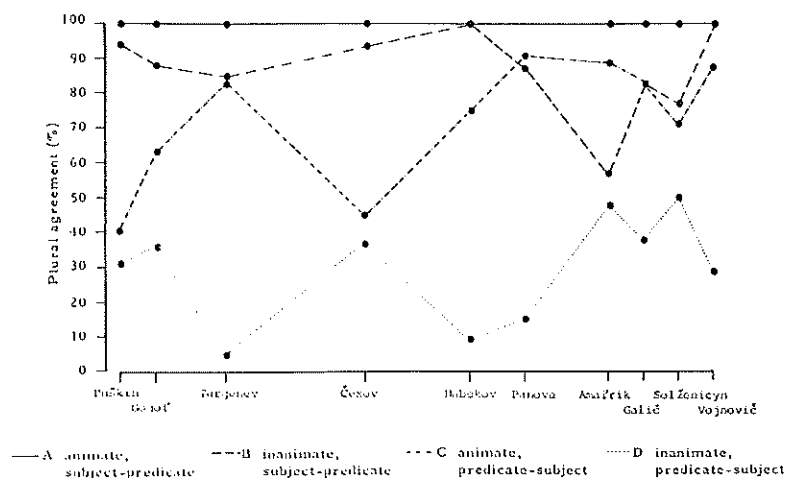


Figure 14: Agreement with conjoined noun phrases in Russian

Here we see a clear pattern. Type A (animate noun phrases preceding the predicate) always scores highest, in terms of plural agreement, with 100%. Type D (inanimates following the predicate) is always lowest. Types B and C come between, but their relative positions can vary. There are striking differences between authors (no two people speak or write in exactly the same

way), but the variation is constrained by the pattern just outlined. Note that this procedure involving the scanning of a corpus is fully replicable.

What then of experiments? A recent project on Russian colour terms has involved experiments, and increasingly since colleagues in Psychology have taken a major part in the research. These experiments do not transfer easily into the lecture theatre, but I wish to mention them because of the collaborative aspect. Surrey has been a good place for working with colleagues and I thank them all, in Computing, Psychology, LIS, HEDERU and Mathematics. I am also tremendously grateful to the people in the Library, Computing Unit, Audio-Visual Aids Unit, the Research Committee and BIL for their contribution to research.²⁵

Let us return to the question of experiments. Traditional experiments form a smaller part of the typical linguist's activity than for those in some other disciplines. Many experiments which linguists would like to carry out are impractical or unethical. But this is not peculiar to Linguistics. Physicists would like to create a star, but while there are grants for star wars there are none for star creation. So we have to look for situations which are close to experimental conditions.²⁶ These conditions may be found in groups of dialects or related languages where there are similar structures, but where slightly different changes occur. In trying to penetrate the mystery of gender systems, particularly the way in which gender develops, we would look to a group of Daly River languages (spoken in North West Australia), which show gender systems in their infancy. As we saw, the different dialects of Andi show how a gender system can expand (a four gender system has been expanded to five and six in different dialects). The Dravidian languages and the Grassfield Bantu languages show the reverse - the contraction of gender systems. And closest to home, the exotic language English has the remnants of a gender system clinging on in *he*, *she* and *it*. This is a story I hope to tell in more detail during the coming months.

²⁵ For help specifically with the preparation of materials for this lecture I am grateful to Mr K. Shaughnessy, Mr I. Clark, Mr J. Derby, Mr S. Heritage, Mr P. Jennings and Ms A. Read (AVA), Mr M. Ashworth (Library), Mr D. O'Brien (LIS) and to Mrs S. Gee (ELI) and Mrs O. Lalor (LIS) for making the recordings.

²⁶ See T. Pateman 'Philosophy of Linguistics' in J. Lyons, R. Coates, M. Deuchar and G. Gazdar (editors) *New Horizons in Linguistics* 2, 1987, London, Penguin, p. 258.

WHY DO RESEARCH IN LINGUISTICS?

Our last question is why do research in Linguistics. Having glimpsed the richness of the world's languages, some may feel the question has largely been answered. I will therefore add just three brief points. First there is a great deal of work going on in Computational Linguistics. The goals include interaction with computers in natural language and machine translation. Colleagues in LIS and Computing have just been awarded an ESPRIT contract of over £400,000 for the Translator's Workbench Project. Such developments are exciting but should not obscure the fact that progress depends on understanding what natural language is.

A second area is speech therapy. Some children, as I mentioned, have difficulties in learning a first language. Speech therapists do remarkable work with them. However, they depend in part on the account of language which linguists provide, and that account needs improving.

Closer to home is the teaching of foreign languages. We teach foreign languages, with some success. But again one of the inputs is an account of natural language and of specific languages, which must also be improved.²⁷

While the value of these applications of work in Linguistics is undeniable, my reason for doing research in Linguistics is to understand. Understanding language is a path towards understanding ourselves.

²⁷ See C. Sanders 'Applied Linguistics' in J. Lyons, R. Coates, M. Deuchar and G. Gazdar (editors) *New Horizons in Linguistics 2*, 1987, London, Penguin, pp. 200-224 for a survey of research on foreign language teaching.

*The following vote of thanks was given by Professor N. V. Smith,
Professor of Linguistics, University College, London.*

Vice-Chancellor, Ladies and Gentlemen.

I'm assured that there are three golden rules for giving a vote of thanks. Unfortunately, no one knows what they are, so I shall have to improvise. I shall at least try to be brief, though I fear I can't be as laconic as the man asked to lecture on "sex" (or was it gender?), who confined himself to "It gives me great pleasure" and sat down.

What gives me great pleasure is to be associated with the inauguration of one of Britain's best and brightest linguists. Grev Corbett is not only an outstanding theoretician, who has made major contributions to the theory of gender, and issues of agreement more generally, but he also satisfies the layman's stereotype of a "linguist" in that he knows lots of languages: and not just common or garden languages, but ones like Serbo-Croat and Sorbian, that most people can't even locate on the map. Almost the first time I came across Grev was when he submitted a research proposal to the SSRC (now the ESRC). One of the spin-offs of his research was to be "A reverse Dictionary of Old Sorbian". We found it quite hard to persuade the finance and economic managers of the Council that this was a commercially viable product... It says much for the quality of his work that he got the money notwithstanding. I was recently reading one of his papers that referred to these languages as well as to Czech and Chichewa, even to Lak and to Nakh - this latter also known as Bats, for reasons I won't go into. It is clear that Grev has no "lack" of "knack".

But as was clear from the lecture we have just been entertained with, G's first love is Russian. I once tried Russian, but it defeated me. Accordingly, I was delighted to discover that it almost defeated Maxim Gorky as well. In the sardonically entitled "My Universities" part of his autobiographical trilogy, he says:

Grammar with all its ugly, narrow rules and ossified formulae, was particularly depressing and I was quite hopeless at squeezing them into a language so alive and so difficult, so capriciously flexible as Russian.

Grev - Professor Corbett as I'll have to learn to call him, of course has not only mastered Russian, but has demonstrated to generations of grateful students as well as us, that grammar isn't quite so ugly or ossified as many have feared.

Even Shakespeare pilloried grammar: Jack Cade in *Henry VI* would have made short shrift of our new Professor, or indeed of our distinguished Vice-Chancellor, if we judge by his words to poor Lord Say:

Thou hast most traitorously corrupted the youth of the realm in erecting a grammar school... It will be proved to thy face that thou hast men about thee that usually talk of a noun and a verb and such abominable words as no Christian ear can endure to hear.

We should be thankful to Vice-Chancellor Kelly, himself an Honourary Fellow of the Institute of Linguists, and to his respected University for the more enlightened attitude that tolerates strong linguists as well as strong solids; and that has made this evening possible and has made it pleasurable - by allowing about him men that talk of a noun and a verb.

Most of all, however, we owe a great debt to Grev Corbett: his erudition and wit have instructed us today, but more importantly we have been given an augury for the future of linguistic scholarship in this University and in this country for the coming years: a good inaugural is a good augury and I invite you to join me in expressing our thanks to Professor Greville Corbett.