The Grammar of Esperanto

Christopher Gledhill

HAL Id: hal-01220008

https://u-paris.hal.science/hal-01220008

Submitted on 28 Jun 2021

ESPERANTO.

A Corpus Based Description.

Contents.

FOREWORD

This book aims to provide a linguistic guide for scholars who are unfamiliar with the Esperanto language. It is not intended as a learner's guide, and numerous teaching materials exist for those who wish to learn the language (Cresswell and Hartley 1992, Janton 1994)[1].

Unlike previous introductions to the language (Wells 1989, Janton 1994) the primary purpose in this book is to set out the social context and grammatical features of Esperanto from the perspective of descriptive linguistics. Despite the unique nature of Esperanto, the phenomenon has been rather neglected in mainstream language studies. Yet there have been many hundreds of dissertations and published papers (for an overview see Tonkin 1997a). As a social movement and an invented language system, Esperanto offers the opportunity to explore in a new light such issues as learner motivation, group identity and the formation of norms and standards in language. For the grammarian, Esperanto offers the chance to observe the evolution of morphology and syntax in a relatively unregulated and highly innovative environment. At different stages, this book attempts to find a response to the following questions:

Is Esperanto a real language?
Which linguistic features are unique to Esperanto?
Is Esperanto a Latin-based language, or some other category?
Does Esperanto evolve, and does this correspond to natural language development?

The grammatical description in this book (section 2 and onwards) follows traditional criteria but adds new information based on a computer-based analysis of authentic texts written in the language. The text archive, or corpus, gives immediate access to the most salient characteristics of the language and its details are set out in the Appendix. Since the first edition of this book, the corpus size has increased to one and a half million words. The corpus is still small by current standards, although this is not surprising given that authentic texts in the language have only recently been gathered together in electronically available collections. Our corpus represents therefore just a broad snapshot of language use. The point of using a corpus is that it provides a representative sample of authentic usage from a variety of texts and from various periods in the development of the language. For example, the corpus allows us to establish how disputed terms or theoretical areas of hesitation may have been

---

[1] The course 'Zagreba Metodo' is available in over 20 languages from *Internacia Kultura Servo* (Amruŝeva 5/1, pf 499, YU-41000, Zagreb). General information on Esperanto is available from the Universala Esperanto Asocio, (Nieuwe Binnenweg 176, NL-3015 BJ Rotterdam).

resolved in everyday use. One point raised by reviewers of the first edition, is that the corpus seems to throw up unusual examples. Most corpus analysts see this as an advantage, in that traditional accounts tend to focus on invented examples which themselves are unlikely to occur in running text and which are usually contrived to the extent that they often miss other more underlying patterns of phraseology (see Sinclair 1991 for a discussion of the principles of 'corpus linguistics'). Researchers of planned or artificial languages such as Esperanto are concerned with the extent to which the grammar as it is used in everyday contexts may deviate from the original plan, or where innovation may have taken the language in a new direction. The corpus allows the descriptive linguist to systematically address some of these questions.

For this second edition of *The Grammar of Esperanto*, I am enormously indebted to Prof. Probal Dashgupta (Hyderabad, India) for suggesting important changes. I would also like to thank Prof. Ian Press (St. Andrews, Scotland) and Mr Liu Haitao (Xining, China) for their advice on Slavic and Asian languages, as well as my wife, Céline and my PhD student Wendy Anderson and the readers of the interlinguistics forum *BJA-listo* who offered comments on the first edition. As usual, any remaining errors of fact and interpretation that remain are my own.

Dr. Christopher Gledhill
St. Andrews University
May 1999

0 BACKGROUND.

Esperanto was published in 1887 by Ludwik Lejzer Zamenhof, an oculist in Białystok (Eastern Poland). The aim was to build a bridge between nations, to allow speakers of different languages to communicate on equal terms in a neutral, user-friendly second language. The original title of the project was *Internacia lingvo* (International Language), although Zamenhof's pseudonym *Doktoro Esperanto* (Dr. Hopeful) became the commonly accepted name for the language. Among the hundreds of similar projects which emerged at the end of the nineteenth century, Esperanto is the only one to have survived to the present day with a sizable number of speakers and a worldwide literature. Esperanto has attracted the support of the League of Nations and Unesco and such figures as Tolstoy, Jules Verne, Gandhi and Einstein, as well as Linguists such as René de Saussure, Mario Pei, Otto Jespersen and Eugen Wüster. But despite significant moral support, Esperanto's political campaign at the level of the European Union and the United Nations has so far enjoyed little success. Instead of claiming that Esperanto is set to become the world's major second language, Esperantists now tend to point out the value of their unique international community (Auld 1976, Tonkin 1997:50). The great ideologies of the nineteenth century, notably Marxism and Darwinism are currently being revisited, and Esperanto in turn is being rehabilitated as a surviving relic of an age of modernist optimism. Over a century after the language's conception, the nomination of one of Esperanto's leading poets (William Auld) for the Nobel Prize for literature is a indication that Esperanto has not yet disappeared from the cultural map.

0.1 Historical development.

The idea of a constructed language designed for international communication dates back to the Enlightenment. The motivation for an invented language was therefore not originally a pacifist ideal, but a genuine attempt to create an international tool for scientific communication. Comenius, Descartes and Leibnitz proposed symbolic systems to represent scientific ideas. Wilkin's 'real character' responded to the need for universal standards in science and represented an attempt to escape the perceived

constraints of natural languages. The resulting symbolic systems of science were later termed *artificial languages* (the term can apply to the terminology of chemistry to the notation of mathematics: Swales 1990). This term was applied later on to universal language projects, although some adherents now prefer the term *planned language* (Blanke 1985). A planned language is a project to develop a new language on the basis of existing natural languages, and whose aim is to create an international medium of communication. The first naturalistic planned languages emerged during the nineteenth century, and since then over 500 projects have been published (Janton 1994:11). The most notable examples include C.K. Ogden's *Basic English,* Otto Jespersen's *Novial* and Giuseppe Peano's *Interlingua*. At the height of the industrial revolution, the universal language movement coincided with an explosion of international communication in the sciences, as well as a rise in international travel and commerce, factors which made the first planned language projects attractive to the new cosmopolitan middle classes (Rasmussen 1997). It was also a time of global imperialism and nationalism, forces which provoked many opponents of these forces to adhere to the new universal language schemes (Tonkin 1997b).

Of the many planned language projects of the late 19th century, Esperanto remains the best known and the most widely used over a century after its first publication. There are a number of reasons for Esperanto's success over and above the rival projects which had preceded it. Compared to other systems, Esperanto's consistent morphology was attractive to learners because it reduced the number of words needed to understand and produce a new text. The 'Sixteen Rules' of the grammar were also alluringly simple. Unlike many inventors, Zamenhof did not insist on instant acceptance of his system: he asked his readers to only learn the language when enough signatures of support had been collected. And to appease language-reformers he offered a period of consultation after which any agreed changes to the language would be incorporated. Zamenhof also appealed to the literary-minded by including translations, original poems and fables alongside the published grammar. Adherents of Esperanto were also attracted by the explicitly ideological stance of Zamenhof's writings. Biographers identify various influences on Zamenhof, not least of which the fact that he was the son of a Jewish censor, living in the multilingual environment of late Tsarist Russia. Zamenhof also had a short-lived attachment to Zionism (Privat 1931:34) which may be related to his later creation *homaranismo*: an anti-nationalist

ideology aimed at uniting humanity under a neutral religion (Tonkin 1997b:73). This pacifist idealism appealed to many at the time, and is still an undercurrent of some Esperantist thinking today (Lo Jacomo 1982). In addition, Jordan (1997) has claimed that Esperanto benefited from the demise of its main predecessor, Volapük, published by a German pastor, Johan Martin Schleyer in 1879. Despite its inscrutable vocabulary, Volapük was during the 1880s a phenomenal success, instantly attracting thousands of followers and backed by hundreds of publications, especially in the sciences. But Schleyer persistently resisted attempts to reform the language and the movement quickly broke up into irreconcilable splinter groups. Esperanto's timely arrival attracted those Volapükists who had argued for a streamlined grammar and recognizable vocabulary. Jordan (1997) suggests that this mass conversion, even if short-lived, reduced the Esperantists' appetite for linguistic schisms. Later, when splits did emerge over reform projects such as *Ido*, Esperanto maintained a loyal following largely because of its existing literary momentum.

Although egalitarian idealism and anti-nationalism have been often cited as reasons for the popularity of late nineteenth century planned language projects, these schemes also coincided with proposals for agreed international standards for measurements, chemical elements as well as postal and telegraphic communications. Even though pacifism and attempts to 'bridge the language barrier' are often identified as the key aims of Esperanto, many early proponents of the language were engineers, technicians and terminologists who were beginning to establish international associations and means of international communication. It is likely that at the time, the legal adoption of an international language of science would have seemed a feasible goal and this coincided with the beginning of the decline of French as the language of education and of scientific publication outside France. In the early decades of Esperanto's existence, an international committee of scientists (the *Délégation pour l'adoption d'une langue auxiliaire internationale*) had been formed to decide which of the many projects would be best suited to scientific communication. The final choice ended in acrimonious disagreement in 1907 when Louis Couturat, the secretary of the Délégation, proposed his own revised version of Esperanto, named *Ido*, a move which was supported by other members of the delegation, including Otto Jespersen. The claims and counter-claims about this incident are surprisingly still alive, as can be witnmessed on the *Ido* sites on the Internet. The episode is remembered as an

important part of Esperanto folklore (*idismo* was the last great linguistic schism in the language) and from a modern perspective serves to demonstrate that at the time, the prospective of a planned language for science was a seriously considered option.

The later history of the Esperanto movement has been documented by Esperantists in some detail (Privat 1931, Drezen 1972, Large 1985, Mattos 1987, Eco 1997). The period immediately after 1908 is generally seen as a turning point. At this time, the *Idists* left the movement, the *Akademio de Esperanto* was elected to control the future development of the language and a more permanent financial footing was found for the *Universala Esperanto Asocio* (UEA) in Paris. From 1908, the infrastructure of the movement was in place, although there were still to be serious disputes over the financial influence of indivduals and different publishing houses (Drezen 1972). Despite the later schisms with *Ido*, any linguistic debate over the form of the language tended to take place in the early period before Zamenhof's death (1917), after which any remaining conflicts took on a more political dimension.

Before the first world war, Esperanto became a household name in Europe and in America, the name itself being appropriated by publicists keen to jump on the bandwagon. As part of the modernist vogue of the time, even brands of cigarettes and soaps as well as café-bars were named after the language. Thousands of clubs and societies formed across Europe and Asia. Esperanto magazines containing essays and poetry were published even for small towns in Scotland such as Arbroath and Dunfermline (from the St. Andrews archives for 1908-1913). However, it seems that the initial social and literary impetus for the movement had vanished during the first world war, perhaps through disillusionment, although Esperantism became a popular pacifist and political movement in the inter-war period. The 1920s saw the formation of ideologically defined pressure groups such as the conscientious objectors (*La Internacio de Militrezistantoj*), produced and published partly by prisoners who had refused to fight in the first war. Despite Esperanto's acceptance by the League of Nations in the 1920s, serious disputes occurred towards the end of the 1930s, especially over the issue of political neutrality and the location of the main offices of UEA. The earlier formation of the socialist *Sennacieca Asocio Tutmonda* (Worldwide non-nationalist association, known as S.A.T.) also reflected the severe political tensions in central Europe. After the second world war the central organization of the

movement was split for a while between different 'central offices' in London and Geneva, and it was only in the 1960s that financial and ideological disputes were resolved and the UEA found a new home in Rotterdam. This schism resulted in the publication of alternative journals and magazines, and seriously weakened the movement. Although the contemporary movement displays considerable political diversity, the last split of note occurred at the Universal Congress 1974 during what became known as the 'Hamburg Putsch', in which the President of the UEA, Ivo Lapenna, and his colleagues were forced to resign over, amongst other things, his policy of 'suppressing communist tendencies'. This part of Esperanto folklore can be seen as a rather minor political event, but nevertheless it importantly marks the beginning of a more stable existence of the movement as a language community rather than as a political entity.

The popular belief that 'Esperanto has had its day' is linked to the movement's phenomenal growth during the first three decades of the twentieth century and its perceived decline thereafter. Esperanto became associated in the public consciousness with failed projects. Surveys carried out by Esperantists often revealed that while people had heard of the language, they also claimed with certainty that 'it had died out' (Large 1985). Ironically, in the British media Esperanto was often compared to the 'pipe dream' of the construction of a channel tunnel linking England and France. However, in the cold war period the movement expanded in Communist countries and became stronger in South America and East Asia. Since the 1960s Esperanto's literature had grown steadily, as had the number of radio broadcasts and other materials in the language. Esperanto is currently still supported strongly by governments (not least by China and the European Community), has became the focus of linguistic research and translation projects (Papegaaij and Schubert 1988) and is supported officially by many non-government organizations (in Italy prominent supporters of Esperanto include the University of San Marino, the automotive firm Fiat and the Italian Radical Democrat Party). Esperanto has been consistently celebrated on thematic stamps by many countries and by local governments in the naming of streets and squares worldwide, while the language can still be seen as an emblematic translation on postcards, in tourist brochures and institutional web sites. In the wider world Esperanto is still recognized partly as a proto-typical example of a

failed project, but also as a symbol of the ideals of peace and international cooperation.

0.2 Sociolinguistic status.

Whereas most invented languages remain projects with few adherents and which often disappear along with their inventors, Esperanto has been adopted by many thousands of speakers and applied to a wide range of cultural activities. Like other speech communities, Esperantism implies a common set of implicit cultural values, means of communication and shared knowledge of cultural artefacts and folklore. But Esperanto is unique in that the motivations for learning Esperanto are entirely different to those of other languages. There is an Esperanto culture, although it is largely abstract and literary. Esperanto culture involves knowledge of writers, personalities and political folklore (Lapenna et al. 1974, Auld, 1986: see also section 4, below). For many, Esperanto is a largely written language, transmitted by the press and international correspondence (in literary magazines, newsletters, academic journals). However, the spoken language has its place in local meetings, and the opportunity for international contacts takes place at weekend retreats and conferences at various locations. Other contacts can be arranged worldwide using the *Delegita Reto* system, which guarantees that representatives (*delegitoj*) will meet the needs of visiting Esperantists. There are also national offices, which have been set up in most developed countries (the permanent office is usually a house or office-block donated to the local Esperantists and serves as the headquarters of the national association). While there are a plethora of radio transmissions from around the world (usually on short wave), sites on the internet and the web have flourished and proven ideally suited to this type of community. The cultural momentum of the Esperanto movement mirrors the minority language movement, and the Esperantists have links with organizations such as the European Bueau for Minority and Lesser-Used Languages (EBMUL, Bañeres and Strubell 1998). The Esperanto movement exploits similar tactics to those used to protect endangered languages such as the Celtic languages (Breton, Gaelic and Welsh) and has parallels with the treatment of Creoles and other newly established languages (Pirlot 1982).

It is impossible to say how many Esperantists there are. Critics who claim that there are 'too few' speakers should bear in mind that speakers of Esperanto certainly outnumber those of rival projects, and for that matter, many endangered and minority regional languages. On the other hand, encyclopedias and propagandists tend to over-estimate the numbers of Esperantists, sometimes quoting millions. A very conservative estimate is around 40,000 speakers, based on the average collective memberships of national organizations affliated to the main Esperantist body *Universala Esperanto Asocio* (this figure is an average taken from 1986 to 1996: UEA, Veuthey 1996). This is a safe estimate because UEA membership is thought to be composed of fluent speakers. However, the figures do not take into account local activists who are not members of the national associations, and cannot account for the many thousands of speakers who were known to be exposed to Esperanto through the education systems of China and Eastern Europe. The volume and frequency of conferences and meetings may also be a reliable indicator of cultural activity: the average attendance at the annual *Universala Kongreso* shows stronger support than for other voluntary international organizations (attendance is on average 2000 for the period 1990-1996). As with many languages, accurate figures for the general population of Esperantists are difficult to find because the definition of 'Esperantist' may range from the fluent 'professional Esperantists' (who work in the national offices) to the least fluent 'perpetual learners' (known as *eternaj komencantoj*). The definition may even be extended to sympathisers at the periphery of the movement who have at best a smattering of the language. Although most usually defined as a second language, it is surprising to find that there are a number of first-language speakers of Esperanto, thanks to the incidence of mixed nationality marriages between fellow Esperantists. In 1974, 200 children from 19 countries were documented and supported by a range of children's learning materials (Lapenna et al. 1974:59, 74). This figure appears to have risen since to 300 families (these are all members of *Rondo Familia* the Esperanto families association reported in Rasič 1994). *Rondo Familia* claims there may be up to 1000 families if non-members are taken into account, and this does not count previous generations of first-language Esperantists. The phenomenon of first-language Esperantists provides important evidence for acquisition of the language and is currently of significant sociolinguistic interest with implications for Creole studies (Versteegh 1993).

Geographically, Esperanto has a wider distribution and a stronger infrastructure than many minority or endangered languages. The UEA contains over 90 affiliated national associations including associations created for autonomous regions such as Flanders and Catalonia. Esperanto is often associated with the regional language movement in Europe, and both have joined forces recently in campaigns to extend minority language rights at a European level. Perhaps ironically, UEA membership is high in wealthy and officially monolingual societies such as Britain, France, Germany, the United States and Japan (Veuthay 1996). While there is no evidence that Esperanto's numbers have declined over the years, the post-war growth of English, the fall in sponsored membership in the UEA (i.e. from former communist regimes such as Poland) and the competition of rival languages may all affect the number of recruits.

Since the beginning, Esperanto's supporters made considerable efforts to construct an alternative cultural heritage. It is even possible to work within an Esperanto-speaking environment (such as the community based around the UEA in Rotterdam and the specialist centres such as the Vienna book archive and the centre for residential language courses at Chaux-le-fonds). Book sales indicate a relatively high demand for Esperanto products, especially for learning materials, reference works and novels. The UEA magazine *Esperanto* reports on average 30 new titles each month and the market is comparable with a regional language like Galician, Occitan and Scottish Gaelic. There are monolingual and bilingual dictionaries, specialized terminological indexes, current affairs magazines, pamphlets, posters, CDs and taped music, CD-Roms, short-wave radio programmes from Eastern Europe and elsewhere, films, plays (performed at conferences), translated and original poetry and novels, graphic novels (including Tintin and Asterix), teaching materials for Esperanto or other languages, scientific and academic journals and special interest newsletters. Although the use of Esperanto in such sciences as cybernetics is often quoted by propagandists, the language has been most extensively used as a vehicle for work in the language sciences, in particular lexicography (Schubert 1993) and terminology (Haupenthal 1978). Of particular note are the range of internet sites available in the language, including literary collections and archive projects. Esperanto's presence in education is carefully documented (Pirlot 1982) and there have been University appointments in Esperanto studies or interlinguistics in the UK, Italy, the Netherlands and elsewhere.

Apart from Esperanto, it is often forgotten that rival language projects also actively campaign in the same 'market' as Esperanto. There are at any one time a number of other alternative projects with similar newsletters and campaigns, and the existence of thousands of sites on the Internet appears to suggest that there is no end of interest in universal languages. Only a handful of perhaps hundreds, notably Glosa (also Glossa), Ido and Interlingua, have been able to attract speakers beyond the lifetime of their original inventor. The main current contender to Esperanto is Interlingua, a reformulation of Peano's project of 1903 and developed by Alexander Gode (1971). Interlingua has enjoyed some exposure at a European level with regular newsletters and meetings. The language is easier to recognize than Esperanto, but does not have the same degree of predictable regularity. There are also differences in objectives and tactics. While Interlingua has some publications and is proposed as a viable project, Esperanto has a community of speakers who are not soley engaged in campaigning for the acceptance of the language. Generally speaking, however no one project has achieved a comparable number of speakers or a similar output of cultural activity as Esperanto.

Sociological research has recently revealed more precise details about who speaks Esperanto, in particular their motivation, age, socio-economic status and ethical profile in various countries (Piron 1989b, Great Britain: Forster 1982 and Large 1985, Canada: Edwards & MacPherson 1987, Germany: Stocker 1996, USA: Jordan 1997). It is often claimed that the average age of Esperantists is high (the movement is frequently described as consisting of senior citizens or pensioners) and thus on the brink of extinction. However, Large (1985) points out that this population is usually replenished by recruits of similar age and Stocker (1996) provides evidence that German Esperantists are younger and more socially diverse than is typically thought, with a roughly equal correspondence of non-professional and young people between Esperantists and the general population. Forster (1982), Edwards and MacPherson (1987) and Jordan (1997) find that there is no strong correlation for any major socio-economic differences between Esperantists and the rest of the population. However, Piron (1989b) suggests that there are clearer tendencies, in particular Esperantists tend to come from the liberal professions (almost three quarters of his sample) as well as a high proportion of administrators and civil servants (Piron used a self-selecting but

very large sample including over 1600 personal adverts and well as details of participants at the annual *Universala Kongreso*). In the Americas, Africa and Asia Piron found a more significant number of blue-collar and manual workers while service and commercial sector employees are significantly underrepresented.

Despite their wide range of socio-economic backgrounds, Esperantists are markedly different in terms of ideology. They may have some social stake in language; for example, they are more likely to be immigrants or language-professionals (Stocker 1996, Jordan 1997). Many Esperantists are translators, language teachers, academics with an interest in linguistics, editors, proof-readers, poets, journalists and the like. Stocker finds that Esperantists are often motivated by intrinsic linguistic curiosity about the language. Younger Esperantists are also more motivated by practical reasons than the older generations and Stocker notes in particular the main reason for Esperantism being the wish to master Esperanto as a third or fourth language (Stocker 1996:50-51). In general the most typical motivation for learning Esperanto is to explore the language as a linguistic system and to access the literature (Auld 1986, Stocker 1996). There is no evidence that Esperantists see the movement or themselves as religious, and Forster (1982) found that Esperantists are more likely to be atheists. Stocker (1996) found that German Esperantists are significantly less religious than the general population, although if they are religious, they tend to belong to non-traditional religions (4% of German Esperantists in the survey follow the *Baha'i* movement).

Most studies emphasize the tendency among Esperantists to be non-conformists. Stocker found German Esperantists four times more likely to be engaged in political activity than other Germans. Stocker's detailed analysis of German Esperantists' value systems suggests that they differ significantly from the general public in terms of their greater positive sense of solidarity, equality of rights and social tolerance. Esperantists differed particularly significantly in terms of their negative attitude towards national identity (Stocker 1996:81-83). This may be a particularly German phenomenon, but it may also correspond with Jordan's observations on American Esperantists. Some Esperantists, Jordan claims, are motivated by dissatisfaction with their own culture and national identity (1997:50). Esperanto thus provides a welcoming and non-threatening alternative culture, where every participant has ownership of the target

language and may even enjoy greater responsibility in this new social grouping than in the wider social community. This form of social escapism can be linked to the fascination with 'toy' languages created for their authors' specific and highly embellished alternative worlds (such as Elfish developed by Tolkien or Klingon from the TV series *Star Trek*). More poignantly, it could be argued that Esperantists share a similar dual cultural identity as Jews or other ethnic or religious minorities (Fettes 1996). The ideological commitment to Esperantism in the early 20th century was often seen as provocative resistance to national authority. Unsurprisingly, Esperantists suffered discrimination and even persecution in the case of Nazism (Lins 1990) and in Stalinist Russia (Piron 1989b:170).

However, there are other motivations for learning planned languages. Esperanto's rival projects such as Interlingua do not claim to provide an alternative culture and they may attract supporters whose motivation for learning is more usually related to a utilitarian philosophy. There is increasing interest in the linguistic analysis of planned languages or *interlinguistics* and linguists often learn Esperanto or other rival planned languages as a matter of professional interest. Similarly, Esperanto scholars have resurrected Volapük and rival 'dialects' such as *Ido* because of their role in the history of the language. One of the main non-ideological motivations cited by learners is the benefit Esperanto might bring in terms of language awareness and linguistic training. Several studies of the role of Esperanto in language learning have been published (including Council of Europe reports and several dissertations, Lapenna et al. 1977). Teaching is seen as a key aspect for the future of the language, with several international associations for teachers such as the *Internacia Esperanto Instituto* who teach the 'Cseh' method, and the UEA-affiliated *ILEI* who set up training courses for teachers and organize international examinations.

The cycle of Esperantism for the individual tends to involve learning the language while relatively young and then returning to actively support the movement at advanced age. The recreational funciton of a foreign language should not be overlooked. While there may be hundreds of people working in Esperanto in the various national offices and in the central office, many Esperantists are simply 'part-time' enthusiasts, attending perhaps a weekly meeting. However, the amateur or 'recreational' nature of Esperantism may not be a major obstacle to the continued use

of the language. As the traditional languages of educational elites decrease in importance (such a trend affected Latin and Greek in the 20th century), and as further education is extended to a wider and older public in many industrialized countries, Esperanto may come to provide an attractive and accessible alternative to learning a natural language.

0.3 The 'Sixteen Rules'.

This section provides a sketch of the grammar published in 1887, officially adopted as the fixed *Fundamento* in 1905. The first eight rules covered the traditional word categories; the second eight dealt with basic phonetic, lexical and syntactic principles. We explore the application of the 'rules' at various points in our discussion of morphology and syntax below.

Rules 1-8: Parts of Speech.

-1 *La* is the definite article. It is invariable. There is no indefinite article.

-2 Nouns end in –*o*. The plural is –*j*. Noun phrase objects require –*n* (accusative case).

-3 Adjectives end in –*a*. They agree for number and case. The comparative is *pli*.

-4 The numerals are *unu, du, tri, kvar, kvin, ses, sep, ok, naŭ, dek* (10), *cent* (100) *mil* (1000).

-5 The pronouns are *mi, vi, li, ŝi, ĝi* (it), *ni, ili.* They agree for case. Possessives end in –*a*.

-6 Verbs only conjugate for tense (past -*is* pres. -*as* fut. -*os*). Participles include -*anta, -ita.*

-7 Lexical adverbs end in –*e*.

-8 Prepositions are used with the nominative (-*o* form). They require no other case.

Rules 9-16: General Principles.

-9 Every word is pronounced as it is spelt.

-10 The accent is placed on the last syllable but one.

-11 Compound words are formed by a junction of roots, the head word at the end.

-12 Negation is expressed by *ne*. There are no double negative expressions.

-13 Direction may be expressed by the accusative -*n* in noun phrases and adverbs.

-14 Where no one preposition is appropriate *je* is used, replaceable by accusative –*n*.

-15 The lexicon may be adapted from any internationally recognizable words.

-16 The final vowel –*o* of nouns and the –*a* of *la* may be replaced by an apostrophe.

Although lexically Esperanto resembles other Latin-based languages, these rules demonstrate a wide variety of influences, including a relatively Germanic and Slavic morphology and innovations such as the explicit signaling of word class. Some rules were very specific in scope, aimed at speakers who might be tempted to use double negatives (as in French) or cases after prepositions (Slavic) or to use some form of indefinite article (English). Others offer useful redundancies: rule 14 allows the speaker to avoid specific prepositions and shifts the role of the accusative to that of 'marked' or object case. Rules 11 and 15 allow for different vocabularies to develop in parallel. And rule 16 was written specifically to allow writers and poets to vary stress patterns, for example, Zamenhof's *Lang' de saĝulo, man' de militisto* 'tongue of the wise, hand of a warrior'. (Kalocsay 1963:127).

Before the 'Sixteen Rules' were published, Zamenhof reformulated his language in a series of projects from 1878 to 1887. A surviving extract of such a proto-Esperanto (*Lingwe uniwersala* adapted from Privat 1931:31) demonstrates that the basic lexicon and morphology were in place, although Zamenhof can be seen to be experimenting with the forms that rival projects, especially *Ido* tried to reintroduce later on (notably plural –s, nouns in –e, verbs in –are, weak vowels such as /j/ in diphthongs):

Lingwe uniwersala:	Esperanto:	English:
Malamikete de las nacjes	*Malamikeco de la nacioj*	Hatred of the nations
Kadó, kadó, jam temp está!	*Falu, falu, jam temp' estas!*	Fall, fall, the time has come!
La tot' homoze in familje	*La tuta homaro en familio*	All humanity in one family
Konunigare so debá.	*Kununuigi sin devas.*	Must as one become.

The issue of control has often been raised at this point: after 1887, how could the invented system remain stable in such a disparate community, all of whose members must have begun as learners? One of the most original aspects of the language is the relative freedom to write and speak without the constraints imposed by a 'native' language community, a point emphasized by Lo Jacomo (1982). However, this literary freedom was not always emphasized in the early years when the ideology of consistency and international standardization prevailed. After several years of debate and correspondence, Zamenhof and his followers agreed a convention in 1905 essentially sanctioning the use of 'unofficial' forms as long as they did not contravene

the Sixteen Rules and the 900 basic words of the first dictionary which were declared 'untouchable'. This convention became known as the *Fundamento*. Esperanto's patterns of usage emerged more fully in Zamenhof's writings and speeches as well as in the many publications by prominent contemporary members of the Esperanto movement who began to experiment with the language. The emphasis on consensus has led to organic developments in the language, since no innovations can be imposed or prevented. Any changes that have taken place since 1905 have involved two broad processes:

Systemic change: re-interpretation or generalization of the rules for new uses (for example, free formation of predicative verbs: *kio gravas* 'what is important', and compound verb forms including tense and aspect markers: *mi manĝantas* 'I am eating', *tio malpermesitas* – 'that is not permitted').

Instantial change: application of general tendencies that are not covered in the rules but have emerged in use (*ĥ* is at times replaced by *k*, while some lexical items have become productive prefixes, such as *ĉef-* 'leading, chief').

The *Akademio,* a body of eminent writers, was set up in 1908 to decide on new (especially technical) terminology and questions of usage, and the evolution of the language was such that by 1970 there were 16 000 basic root words (Wells 1989). In terms of grammar, influential monolingual works such as the *Plena Analiza Gramatiko* (or PAG, Kalocsay and Waringhien 1985) and the *Plena Ilustrita Vortaro* (or *PIV*, Waringhien 1970 with supplements to 1987) have served to fix certain usages for advanced learners, although there has always been debate over a number of conventions. The passive *'ata / ita'* distinction is perhaps the most notorious and is also discussed below. Despite the official tone and size of these dictionaries, many Esperantists refuse to accept certain 'innovations' and point to popular usage, or to analogies which seem to disprove the 'official' recommendations. Other changes have been widely accepted without official sanction, or have favoured one 'official' source over another. For example, according to the *PIV* the term for 'television' is *televizio* but all Esperantists in fact use the term accepted by the *Akademio*: *televido* (*tele* + *vidi*: to see). In other cases, 'local' Esperantists refuse to accept usages introduced by the 'professional Esperantists' who work for the UEA (this accounts for varying uses

of the words such as 'computer' *komputilo, komputoro, komputero* or 'software' *softvaro, programaĵo*). Magazines and local clubs still debate the correct use of the language and readers check each others' speech and written production with a wide range of dictionaries, grammars and phrase-books. As we have noted above, Esperantists are above-average linguists but also in the main non-conformists, and this engenders considerable debate on language issues within the movement. We should not emphasize variation too much however. National and bilingual dictionaries as well as the teaching materials which have been developed in the various ethnic languages are perhaps the most influential disseminators of 'Standard Esperanto'. Despite the existence of alternative dictionaries (Cherpillod 1988) and dictionaries of slang (Corsetti 1987), most teaching materials respect the basic grammar and vocabulary set out in the *Fundamento* and this consensus accounts for the degree of homogeneity that does exist in the language as a whole.

1. PHONETICS AND PHONOLOGY.

Table 1: The Alphabet.

a	/a/	b	/b/	c	/ts/	ĉ	/tʃ/
d	/d/	e	/e/	f	/f/	g	/g/
ĝ	/dʒ/	h	/h/	ĥ	/x/	i	/i/
j	/j/	ĵ	/ʒ/	k	/k/	l	/l/
m	/m/	n	/n/	o	/o	p	/p/
r	/r/	s	/s/	ŝ	/ʃ/	t	/t/
u	/u/	ŭ	/w/	v	/v/	z	/z/

Table 2: The Phonemes of Esperanto.

Vowels	Front	Mid-front	Central	Mid-back	Back
High	/i/ spread				/o/ rounded
Mid		/e/ spread		/u/ rounded	
Low			/a/ open		

Consonants	Bilabial	Labio-dental	Dental	Post-alveolar	Palatal	Velar	Glottal
Stop (-voice)	/p/		/t/			/k/	
Stop (+voice)	/b/		/d/			/g/	
Affricate (-voice)			/s/	/tʃ/			
Affricate (+voice)			/z/	/dʒ/			
Fricative (-voice)		/f/	/ts/	/ʃ/		/x/	/h/
Fricative (+voice)		/v/		/ʒ/			
Nasal	/m/		/n/				
Lateral				/l/			
Uvular roll				/r/			
Semi-vowels	/w/				/j/		

1.1 Lexicon.

Esperanto's vocabulary is usually described as a mix of words from Latin, French other European languages. The picture is much more complicated than this, however. Zamenhof also equipped the language with a stock of grammatical words which resembled no one language in particular. More importantly, despite originating from one or another language, the majority of words in Esperanto were modified and accommodated into a more neutral form (for example, *ŝati* 'to appreciate' comes from German 'schatzen'). In addition, the basic morphological rules of Esperanto inevitably led to word forms that were unique to the language (below, both categories are classified as 'Esperanto' words). The following table sets out the origins of the first 1000 most frequent words in our corpus and this reveals a much wider influence of the home-grown vocabulary (details of the corpus are set out in the Appendix):

Table 3. The Origin of Esperanto Words.

	Latinate	Esperanto	Germanic	Indo-European	Greek	Balto-Slavic	Total
First 100 words	47	26	7	17	1	2	100
First 1000 words	704	121	101	50	15	9	1000
	70%	**12%**	**10%**	**5%**	**< 2%**	**< 1%**	**100%**

The first ten most frequent words for each language category are:

Latinate: *la* (the), *de* (of), *en* (in), *por* (for), *ke* (that, conj.), *sed* (but), *kun* (with), *ankaŭ* (also), *per* (by, with), *sur* (on).

Esperanto: *kiu* (who), *tiu* (this one), *tio* (that PRO), *kiel* (how), *kiuj* (who PL), *kiam* (when), *tion* (that ACC), *ĉiuj* (everyone PL), *tiuj* (those PL), *tiel* (thus).

Germanic: *nur* (only), *ŝi* (she), *havas* (has), *ja* (indeed), *jen* (here is), *landoj* (lands), *sendu* (send), *dankon* (thanks), *daŭre* (continues), *suno* (sun).

Indo-European: *al* (to), *mi* (me), *estas* (is), *ne* (no), *li* (he), *ni* (we), *el* (out of), *oni* (one), *da* (of *quantity*), *unu* (one *numeral*).

Greek: *kaj* (and), *eŭropo* (Europe), *ekonomio* (economy), *celo* (aim), *teknika* (technical), *temo* (theme), *ege* (greatly), *sfero* (sphere), *tekniko* (technique), *celojn* (aims ACC, PL).

Balto-Slavic: *ĝi* (about), *ol* (than), *ĉu* (whether, *interrogative particle*), *tuj* (immediately), *nu* (well, *interjection*), *krom* (besides), *nepre* (certainly, necessarily), *pravis* (was right), *po* (per, at the rate of). [only nine words within the first 1000].

Several figures have been advanced to describe the distribution of Esperanto's vocabulary. Janton (1994) suggested that the origin Esperanto's vocabulary is divided as follows: Romance (Latin / French) 80%, Anglo-German 10%, Greek 5%, Slavic 5%. While these calculations are based on dictionary head-words and privilege the range of vocabulary in the language, they also tend to underestimate the role of grammatical items which are only counted once in the dictionary, as well as compound words, which are sometimes not included in the dictionary at all. In any language, the most frequent 100 words typically make up around 60% of any text, while the most frequent 1000 words will typically account for 85% of a text (Sinclair 1991). The most frequent 1000 words are therefore the core words in the language, the central vocabulary that can be used in a variety of contexts and combinations. Thus the corpus-based count presented above gives us a picture of typical language use rather than a picture of potential and (at times) peripheral use. As can be seen in table 3, speakers of Esperanto use a somewhat smaller distribution of Latinate words than suggested by Janton, and the situation is complicated by a larger number of Germanic forms and by very frequent use of words derived from Esperanto's own morphological system. Grammatical items in particular are original or derived from common European roots rather than Latin, a situation which makes Esperanto similar to English in this respect: English is grammatically a Germanic language but is also

described as 'latinate' in terms of its vocabulary. The following notes explain the distribution of Esperanto's word stock (they are presented in order of frequency):

Latinate words (70%) dominate the core vocabulary of the language. A small but frequent group of words are derived from classical Latin forms and bear little resemblance to modern Romance languages (*sed* but, *dum* during, *apud* next to, *tamen* however, *hodiaŭ* today, *kvankam* although). The majority of other Latinate forms are lexical items, corresponding to Italian words which resemble cognate words in French, Spanish and Portuguese (*aŭtuno* autumn, *fari* to do, *granda* large, *legi* to read, *popolo* folk). A broader but less frequent range of lexical items correspond to the 'Renaissance' vocabulary which predominates in English and in most other European languages: concepts derived from Latin and propagated by Italian and French in the 17th, 18th and 19th centuries (*literaturo* literature, *industrio* industry, *politiko* politics, *revolucio* revolution, *scienco* science). The 'Latinate' category also includes a number of high frequency words and grammatical adverbs from French which have been heavily modified and do not resemble other Romance languages: *mem* emphatic self, *plu* further, *tre* very, *apenaŭ* hardly, *malgraŭ* despite, *preskaŭ* nearly. However, adverbs in -*aŭ* are not exclusively derived from French as can be seen with *ankaŭ* ('also', Italian), *ankoraŭ* ('again', Italian), *almenaŭ* ('at least', Spanish).

Zamenhof also chose to create a small number of words which do not resemble any one particular form, although they may have a passing resemblance to Romance-based words. This includes the very common grammatical items *eĉ* (the specifier 'even' is, derived from the Italian adverbial use of 'eccetto' 'except, besides' although Esperantists also claim that it is from Latin 'etiam'), *laŭ* ('along, according to' perhaps from the Italian 'lungo' used as an adverb or preposition), *ĉi* proximative 'this' (resembling French 'ci'), *ĉe* ('at', from French 'chez') and *ĝis* ('until, up to' possibly from French 'jusque'). These words are at the very edge of the 'Latinate' category however, as their use differs considerably from the original languages and their modified nature leads us to consider them to be 'Esperanto' words in their own right.

Esperanto words (12%) are unique to Esperanto and do not resemble any ethnic language forms. Detractors claim that this puts Esperanto at a disadvantage compared with projects which use better-known Latinate forms (notably, Interlingua), while

Esperantists point out that the core words of Esperanto represent a clear break with European languages, and that in any case even very closely related languages are often multually unintelligible. Most of the very high frequency 'Esperanto' words are grammatical items corresponding to the family of words known as correlatives. This includes common pronouns (*tio* that) and grammatical adverbs (*kial* why) (see our discussion of Indo-European words below and chapter 2.4). Although the principle of the correlatives can ultimately be derived from Lithuanian, the words are unrecognizable for non-Esperantists and represent an important proportion of texts in the language.

Outside the first 100 words, unique 'invented' forms tend to be rare and are usually words created by Zamenhof or other Esperantists on the basis of an analogy (*aboco* alphabet, *usono* the U.S) or for stylistic effect (*licas* 'it is licit, permitted'). The word *edzo* 'husband' was thought to be the only truly invented word in the language with no foreign analogy, although it has been claimed that *edzo* is regularly derived from *edzino* 'wife' on the basis of a German title once used in Poland *Prinzedsin* ('Prince's consort'). More importantly, the 'Esperanto' category includes compound words, of which around 25 occur in the most frequent 1000 of our corpus. Although these may be ultimately derived from European languages, they represent combinations unique to Esperanto (*ekde* as from, *junularo* youth movement, *malpli* less, *iomete* somewhat, *preterpasi* to take over, to pass). Since the average number of derivations of a single lexical item is around ten, this leads to a large number of low-frequency derived forms which will not appear within the first 1000 words. Moreover, the frequency of compound words rises as the overall vocabulary increases (thus compound words are already more frequent than Germanic words in the most frequent 1000 words, and they begin to catch up with Latinate words if we consider the most frequent 10 000 words). These forms are usually 'hapax legomena' (one-off constructs) and apart from the frequent forms we mentioned above, they rarely find their way into standard dictionaries. Compound forms are discussed under morphology (chapter 2).

Germanic words (10%) are infrequent in the first 100 words, but a relatively large number appear in the first 1000. Germanic words therefore represent an important core of the basic vocabulary. Zamenhof selected items from German and Dutch (*anstataŭ* instead of, *baldaŭ* soon, *frua* early, *nur* only, *tago* day, *taŭgi* to suit, *ŝajni*

to seem, *vorto* word) as well as English (*boato* boat, *jes* yes, *ĵus* the adverb 'just', *sendis* sent, *taskoj* tasks). Many German forms are slightly changed or have affinities with Scandinavian words (for example Swedish 'tinga' = *dungi* to employ). The comparative expression *ju...des...* as in *ju pli frue des pli bone* (the earlier the better) also comes directly from Swedish ('ju…. desto'). German words are particularly common in *ŝ*-based consonant clusters (*ŝlosi* to lock, *ŝminko* make up, *ŝtormo* storm, *ŝpari* to save, *ŝtato* state). English words are less frequent in the basic word stock, but the influence of the language has led to the introduction of new words throughout the 20th century: *basketbalo* basketball, *ĉipa* inexpensive, cheap, *kompilero* compiler, *softvaro* software, *dezajno* design (alternatively spelt as *dizajno*). The Germanic languages may have indirectly inspired two forms in particular *je* (an important neutral preposition in Esperanto, perhaps derived from the German particle used in idiomatic phrases as in 'je nachdem', 'according to circumstances') and *ajn* indeterminate 'any' (possibly a reformulation of English 'any' or the first syllable from the German 'einige' – although 'irgend' is the usual equivalent in German).

'Indo-European words' (5%) include a number of items which represent a compromise common to Latin, Germanic and Slavic languages (such as *mi* me, *ni* us, *vi* you, *si* self, *unu* one, *dek* ten). The Indo-European category seems appropriate for a number of Zamenhof's selections. For example, some Latinate words have several cognate forms in other languages and appear to have a general Indo-European form (*al* = 'to, towards', as in 'alight' from Latin 'ad', *el* = 'out of, from', as in 'elaborate' from Latin 'ex' or 'ab'). Generally speaking we have included in this category only pronouns, numerals and a small number of prepositions which resemble Latin but are widely shared by other Indo-European languages. Some Indo-European or Latin words in Esperanto are so widespread that they correspond to forms in Celtic and Indo-Aryan languages as well (in Welsh *mi* = mi 'me', *ni* = ni ' us', *veziko* = chwysigen 'blister', *kredi* = credu 'believe', *libro* = llyfr 'book', *naŭ* = nau 'nine', and in Sanskrit *estas* = asti 'is', in Hindi *mi* = /mɛ/ 'I', *nokto* = /naktam/ 'night', *du* = /do/ 'two', *naŭ* = /nau/ 'nine'). The correlative family of words is also close to the Indo-European pattern. While the endings are unique to Esperanto, the correlative prefixes *ki-, ti-, neni-* resemble forms common to many European languages, in particular Lithuanian (a Balto-Slavic language) as well as Indian and other Slavic languages. For example

ki- corresponds to English or Germanic w- as in *kio* 'w̲ho, w̲as', *kiam* 'w̲hen, w̲ann', both of which are 'k̲oji, k̲ada' in Croatian or 'k̲yā, k̲ahā̃' in Hindi, while *ti-* corresponds to th/d- in English and German as in *tio* 'th̲at, d̲as', *tiam* 'th̲en, d̲ann', and in turn 't̲o, t̲ada' in Croatian. In Hindi only /t̲ab/ correspond to *tiam* /'then' as the k̲ forms are highly consistent in Hindi. The t̲-forms are apparently more consistent in the related language, Marathi (Dashgupta, personal communication).

Greek words (< 2%) are a largely negligible part of the basic word stock of the language, although they are more frequent in literary or technical discourse (accounting for the high score for Greek words cited by Janton). The word *kaj* 'and' represents the clearest example of a Greek borrowing, although it does not correspond to the modern Greek pronunciation, which is /kɛ/. Most of the examples we cite above have been transmitted as French and Latin renaissance forms (*temo, ekonomio, sfero*) and as such they can justifiably be classified as Renaissance Latinate, while *celi* 'to aim' can equally be seen to be derived from German 'zielen' (ultimately from Greek /telos/). Nevertheless, many lower-frequency Greek words are core items in Esperanto (for example, they occupy the majority of the *h-* section in the dictionary: *heliko* snail, *heroo* hero, *hipoteko* mortgage) and they have had a considerable influence on European technical morphology which is represented in forms such as *arĥi* (archi-), *hidro* (hydro-), *proto* (proto-) and so on. The influence of Greek can also be seen in the formation of the plural (in Greek κατραφρακτωζ 'kataphraktos' becomes κατραφρακτωι 'kataphraktoi' pl. 'cavalrymen'). Zamenhof intended 's' to be the plural, but preferred instead the Greek semi-vowel /j/, as it permitted the addition of the accusative morpheme *–n* with no subsequent change in the word form (the original choice was to have a vowel change for the accusative case with *–s* plural). Finally, it is claimed that the widely used morpheme *-eg-* (augmentative, as in Italian -issimo) was derived from Greek 'mega' and this form has found its way into normal usage as a free morpheme (*ege* greatly).

Balto-Slavic words (< 1%) are under-represented in Esperanto, although many of the 'Indo-European' forms we considered above are more widespread in Slavic than in the Romance or Germanic languages (*vi* you, *ne* not, *domo* house, *raketo* rocket, *sidi* to sit, *vidi* to see). Among the most frequently-used Slavic words in Esperanto,

Zamenhof selected specific items from Russian (*nu* 'well' interjection, *krom* 'besides' from 'krome'), Serbian / Croatian (*ol* 'than' corresponds to 'od') and Polish (the interrogative particle *ĉu* is from 'czy', although 'czy' is only used clause-initially) as well as general Slavic forms (*prava* true, correct, *ĉerpi* to extract, *kajuto* hut, *klopodi* to strive, *kolbaso* sausage). Zamenhof was also familiar with Lithuanian (a Balto-Slavic language), as can be seen in his choice of *po* (at the rate of, but also meaning 'under' in Lithuanian), *tuj* ('immediately', corresponding to 'tuojau') and *ĝi* (a neutral pronoun 'it', corresponding to 'ji' the feminine pronoun for 'she, it' and contrasting with 'jis' the masculine 'he, it'). Although the correlative forms are unique to Esperanto, they can also be seen to have been inspired by Lithuanian (see section 2.4).

Some words appear to be highly modified forms of Slavic words (for example *nepre* 'certainly' is derived from Russian 'nepremenno'). Words such as *pri* 'about, concerning' and *pro* 'because of, due to' appear Slavic, but have different meanings to their original form (in Russian 'pri' is less specific and means 'on, around' while 'pro' conversely signifies 'about, concerning'). Finally, a general Slavic influence on the language appears in the distinction between perfective and progressive verb participles (realized by various vowel changes in Lithuanian and the Slavi languages and as *ata / ita* in Esperanto) and the infinitive verb ending *–i* (often in the form '=i', '-eti' or '-yti/s' as in *sidi* (to sit) which is 'sediti' in Croatian, or in Lithuanian 'sėdėti').

1.2 Orthography.

The phonology of Esperanto was set out in the *Fundamento*. The alphabet consists of twenty-eight letters, all of which correspond in principle to one phoneme sound (i.e. rule 9: every word is pronounced as it is spelt). The IPA phonetic equivalents and phonetic properties are given in the table on page 16. The Esperanto alphabet is read out aloud phonetically with *–o* added to the consonants (hence ABC is pronounced /a-bo-tso/). The letters <q, w, x, y> are not productive in word formation but are pronounced /kuo, vavo, ikso, ipsilono/. To give some idea of the distribution of letters and sounds in the language, in Appendix 2 we have set out an analysis of the distribution of letters. The corpus is a written archive, and typically has higher

numbers of proper nouns (i.e. in journalistic accounts) and lower numbers of technical or scholarly words (which may contain $\hat{h}$ more frequently). Nevertheless, the statistics suggest that *a, d, e, k, l, s* represent the most widespread letters (each with more than 20 000 occurrences). In contrast the letters *ĝ, h, ĥ, ŝ, u, ŭ, v, and z* all occur less than 5 000 times in the sample. *U* is the least common vowel, while *ŭ* is infrequent because it occurs almost exclusively word-finally (except in high frequency words like *Eŭropo*). *J* and *ĵ* are also more likely to be involved in word-final morphology. The letters *ĝ, h, ĥ, ŝ, v* and *z* are however generally under-represented, especially in mid-word or syllable-final positions. The letter $\hat{h}$ is considered as particularly marginal. It originally represented 'chi' from Greek words and is still used to transliterate names. In many cases it has been replaced by another consonant such as *k* as in many European languages. In common words, $\hat{h}$ survives in alternation with *k*. From the corpus, it appears that the letter is in decline but has not disappeared. The form *teĥn-* (as in *teĥnologio*) is its most frequent use (although even here the form *teknologio* is more frequently used). The letter is used interchangeably with *k* in the words *ĥaoso / kaoso* 'chaos', *ĥemio / kemio*, 'chemistry'. A very small number of words retain $\hat{h}$ however in order to avoid homonyms: thus *ĥolero* 'cholera' is distinguished from *kolero* 'anger', as is *ĥoro* 'choir' from *koro* 'heart'.

Conventions for acronyms, hyphens and punctuation are not strict. Generally speaking however, writers are encouraged to use commas according to East European or German conventions rather than English or French (that is commas are used to introduce noun-phrase complement clauses such as *La ideo ke,* The idea that… and relative clauses *La viro, kiu…* The man who….). Writers tend to use hyphens to separate the longer word compounds rather than syllables (and to avoid confusion with double letters formed by compound words, for example 'ch' in *senc-hava*, meaningful). In compound words, the final grammatical morpheme is usually deleted especially when the first letter of the next item is a vowel (*laborkondiĉoj = laboro-kondiĉoj* working conditions, *falakvo = falanta akvo* falling water, a shower) although this is not obligatory (*grandanombre = je granda nombro =* in great numbers, *akvofalo = falo de akvo* waterfall). Additionally, in the *Fundamento* Zamenhof demonstrated the morphology of the language to learners by separating them graphically (thus *Esperanto* would be written *Esper'ant'o*).

The explicitly mixed nature of Esperanto has resulted in a number of compromises in an attempt to find international spellings. In addition, the language was deliberately cleared of homonyms leading to changes in the form of certain words. Esperanto's lexicon was aimed at the cultural elite of central Europe who could be expected to be familiar with Latin, Greek, French and German and the accented roman alphabet of the central European languages. In Zamenhof's original design and in its subsequent development, several sometimes incompatible processes affected the form of Esperanto's words. We set out the spelling conventions of Esperanto here according to two divergent principles, orthographic consistency (the need to create a language with a coherent spelling system) and etymological transparency (the need to create a language with recognizable words). All etymologies here are attested in Bastien (1950).

Orthographic consistency:

- To obey the general morphology of the language, borrowed words received grammatical endings, a process which necessarily interrupted the original spelling (*-o, -a, -i* for lexical items). In addition, since the phonologies of most languages (especially English and French) have outgrown their spelling systems, borrowed words in Esperanto are almost always phonemically respelt. Many Italian words were chosen because of their simple Consonant Vowel (CV) structure, but they similarly undergo some change: *ĉielo* from 'cielo' comparable with 'caelum, ciel, sky', *kapo* from 'capo' vs. 'cabeza, head, tête'.

- Silent letters and diphthongs were generally simplified: *brako* 'bracchium, bras, braccio' (arm), *bezono* 'besoin, bisogno' (need), *apogi* 'appuyer, appoggiare, apoyar, apoiar' (to support: note the hard stop sound in /g/). In many words however, consonant clusters were often maintained (*sandviĉo*: sandwich, *strato*: stratum, strada, a street, *ŝrumpi*: einschrumpfen, to shrink). In other cases diphthongs and vowel clusters corresponded with original spelling rather than with the pronunciation in the donor languages (*grajno* grain, *trajno* train, *boato* boat).

- The desire to avoid homonyms resulted in the formation of a large number of deliberate paronyms: words historically derived from the same root. For example, *pesi* (transitive) and *pezi* (intransitive) represent both meanings of 'to weigh', *cendo* is

chosen for 'a cent' to avoid '*cento'* (a hundred of…), *magazeno* 'shop' is distinguished from *magazino* 'magazine', *generalo* 'a general' vs. *ĝenerala* 'general' (adj.), *poento* 'point, score' vs. *pinto* 'point, peak'. A small number of homonyms were later created however, mainly through the development of technical words. Wells (1969) lists *vato* (1) cotton wool, *vato* (2) a watt, *cisto* (1) a cyst, *cisto* (2) a rock rose (in botany). In some cases, the second item undergoes doubling of the last root vowel or consonant to provide a different spelling: thus *vatto* (a watt of electricity)*, Finno* (a Finn). In these cases therefore Esperanto displays 'true' consonant gemination (as in Italian) as opposed to gemination caused by morphological clustering.

Etymological transparency:

- Similar terms in the European languages were chosen, usually following the most widespread pronunciation: *rifo* (reef) from 'récif, arrecife, recife, Riff, rif', *ĉevalo* (horse) from 'caballus, cheval, caballo', *fari* (to do, make) from 'facere, faire, fare, hacer, fazer'. *zibelo* (sable) from 'zibeline, zibellino, cebellina, zebelina, Zobel, sobol'. *enui* (annoy) from 'ennuyer, annoiarsi, enojo', *situacio* from 'situation, situazione, situaçiaĝ, situación'.

- The original Latin (and sometimes Greek) form was sometimes preferred to the modern European term, a process which often simplified pronunciation, e.g.: *koro* from Latin *cor* vs 'coeur, cuore, corazón', *ordo* from Latin 'ordo' instead of 'ordine, ordre, Ordnung, order', *urbo* from 'urbs' instead of a more 'modern' form such as 'ville, Stadt, civitas, city, cité, ciudad, città'. But a derived form was preferred where it was likely to be recognized than the original form: *leĝo* (law) from 'légal, legge' rather than Latin 'lex', *bovo* (cow) from 'bovin, bovine' rather than 'bos'.

- The importance of French in Europe in the nineteenth century meant that its words were still widely recognized and preferred despite the relative simplicity or widespread nature of other forms. Thus *emajlo* (enamel, 'émail') is preferred over 'smalto, esmalte, emal', *ĉerko* (cercueil 'coffin') over 'sarcophagus, sarcofago, Sarg, sarkofag', *kuiri* (cuire) over 'coquere, cucinare, cocer, cozer, to cook, kochen'. In other cases, however, a Germanic (often English) form is preferred (*birdo* instead of oiseau, ave, uccello, Vogel, and similarly for *rusto* for rust, *ŝipo* for ship). A smaller number of purely Slavic terms were preferred although usually in a modified form (*ĉerpi* to

draw out, *klopodi* to strive, *nepre* undoubtedly). It has been suggested that Esperantists preferred the early Slavic forms because they made the language less Latinate and more 'international', while the lexicographer and grammarian Waringhien was often criticized for deliberately choosing French-based roots. The early influence of French and the later influx of English terms (such as *softvaro* software, *mitingo* meeting, *manaĝero* manager) reflects the usual tendency to borrow from currently dominant languages. On the other hand, some of these 'borrowings' are in fact simply lexicographic conventions; and the English based *manaĝero* was accepted into later editions of the *Plena Ilustrita Vortaro,* but was not in fact a widespread term used in the community. Similarly a small number of words introduced by Zamenhof (such as *fakelo* 'torch') were ignored as other terms became prevalent in the community (the word French and English-based *torĉo* is preferred).

As we pointed out in the previous section, Esperanto's vocabulary is an eclectic mix, and there are clearly several competing principles in operation at the same time. Esperantists defend the overall shape of the lexicon by appealing to Zamenhof's sense of esthetics; he was after all more interested in composing written texts than accumulating etymologically consistent sets of words. By attempting to provide a monomorphemic system with as few homonyms and homographs as possible, Zamenhof and later Esperantists had succeeded in providing a streamlined lexicon which if not consistent, at least obeyed the considerable constraints of an agreed grammar and derivational morphology. Although critics point to the fact that much of the vocabulary is derived from a few ethnic languages, especially French and Latin, we see below that the forms of these words are usually radically altered by morphology.

Apart from the issue of recognizabiliy, the alphabet has also been the source of criticism and much Esperanto mythology. The accented letters (*ĉ, ĝ, ĥ, ĵ, ŝ, ŭ*) drew particular criticism because they only corresponded to the typefaces of a few national scripts. One story has it that in the early years a number of publishing firms were interested in obtaining exclusive contracts with the major Esperanto journals. Drezen (1972) claims that once the publishers Hachette had won this contract, they installed the accented letters at the typesetters and were unwilling to allow them to be adapted for competitors. Zamenhof defended the use of accents in order to maintain the one-

letter / one-sound relationship, although he suggested that they could be replaced by digraphs of the same letter + *h* (*ch, gh, hh, jh, sh*, where *ŭ* would be left as *u* or some marked variant, e.g. *ú*). However, nowadays the issue is of less relevance. Word processing has largely overcome the problem, although the *h*-remedy is still occasionally seen (it is used in our Appendix). Recently, *x* has come to be used instead of *h* in electronic mail and on the internet. More fundamentally, in recent years there has been considerable debate in the Esperanto press about whether national words and proper nouns should be transliterated or left in their original form. Purists argue that foreign forms debase a perceived 'logic' of the system while others claim that Esperanto should respect original cultures, especially non-western and non-European spellings. The general solution in the early 20th century was to agree common Esperanto forms in an attempt to provide a rationalized system (*Novjorko* New York, *Vaterlo* Waterloo, *Ŝekspiro* Shakespeare, *Darvino* Darwin) but towards the end of the century the tendency has been to leave names intact (or standard transliterations in the case of non-roman script names, as in *La vivo de Gandhi* 'The Life of Gandhi').

1.3 Pronunciation.

No standard 'accent' has been written into the design of Esperanto. Zamenhof (1962) simply gave equivalent pronunciations of each letter in a variety of European languages and stated that syllables and phonemes should be pronounced with as little variation as possible. In terms of quality (but not in terms of length and other features) the five vowel system clearly emulated Spanish and Greek and these speakers are often referred to as models for pronunciation. Wells (1989:26) cites Serbo-Croatian and Czech speakers as suitable models, while Kalocsay (1963:119) claimed that Italian should provide the model intonation pattern for Esperanto (although he presumably did not want to extend this to prosody at sentence level). Recordings by prominent Esperantists are available for reference and various spoken materials exist for teaching. Vowel length, nasality and tone are not meaningful elements of the phonology. According to the one-sound / one-symbol principle, diphthongs are only formed officially with the semi-consonants *j* /j/ and *ŭ* /w/. The only commonly occurring combinations of this type are /aw, ew/ and /aj, ej, oj, uj/. Their use is fairly

restricted although *-aŭ* is a common neutral ending, and *–aj, -oj -uj* are very common plural forms. Other vowel combinations are common but are treated as bi-syllabic although diphthongs can also be heard in casual speech. Officially, for example there should be no semi-vowel in *tradicio* /traditsio/, although it is sometimes pronounced with an intervening diphthong: /traditsijo/. Bi-vocalic combinations typically occur in the correlatives such as *kiel* (how, as), *iom* (somewhat). Bi-vocalic combinations also occur in derived lexical words: *treege* (very much), *praulo* (ancestor) where some speakers may erroneously introduce /j/ or /w/ between the vowels. Vowel combinations also occur in single lexical items and in paronyms (i.e. forms designed to avoid homonyms: *firma* firm (adj.), *firmao* a firm, *metro* one metre, *metroo* the metro, *buro* a Boer, *buroo* a bureau, *post* after, *posteo* posterity).

Unsurprisingly, most of the consonantal distinguishing traits of European languages are present in Esperanto, including voice, friction, nasality. Other features are discouraged, however. Assimilation in consonant clusters is seen as a transgression of the phonetic principle and is discouraged (Wells 1989:24): the /k/ in *ekzameno* should remain unvoiced and not be pronounced */egzameno/. Similarly, vowel harmonization and syllable-reduction are discouraged. Some Slavic consonant clusters prove to be difficult for learners (as mentioned below). More frequently, consonant clusters are formed by derivational compounds and agglutination. Difficulties caused by word compounds can be avoided with the introduction of word class morphemes (always vowels). Thus *skrib̲o̲tablo* (writing-desk) is often preferred to *skribtablo*. Some compound forms produce difficult consonant clusters but tend not to be remedied by an intervening vowel. These tend to involve polysyllabic prefixing elements as in *vaporŝipo* 'steam ship' or *matenmanĝo* 'breakfast' (morning-meal). In these compounds a pretonic accent is sometimes maintained on the first syllable (as though the lexical item was isolated) as well as the 'official' penultimate syllable of the entire compound while a slight glottal stop or pause is sometimes heard between the main elements: /vápor'ʃípo/, /máten'mándʒo/.

Consonant clusters in Esperanto are set out below. If only single morpheme words are taken into account, the following pattern emerges for syllable-initial forms (typically

the beginning of words but also in mid-word position) and syllable-final forms (clusters which only occur in the middle of words):

Syllable-initial and syllable-final clusters:

blinda (blind), *breto* (shelf)
drato (wire)
flugo (flight), *fromaĝo* (cheese), *ftizo* (consumption)
glata (smooth), *gnomo* (dwarf), *greno* (grain), *gvati* (watch)
klaĉo (chat), *krom* (besides), *ksifio* (swordfish), *kvar* (four)
mnemoniko (mnemonic)
plaĉi (please), *pneŭmatiko* (tyre), *pruvi* (prove), *psiko* (psyche), *pterido* (bracken),
sceno (scene), *sfero* (sphere), *skalo* (scale), *sklavo* (slave), *skvamo* (scale), *skribi* (write), *slavo* (Slav), *smeraldo* (emerald), *snufi* (sniff), *sporto* (sport), *splena* (depressed), *sprita* (witty), *stacio* (station), *strato* (street), *svingi* (swing)
ŝkoto (sheet sail) *ŝlifi* (polish), *ŝmiri* (smear) , *ŝnuro* (rope), *ŝpari* (save), *ŝpruci* (spurt), *ŝrumpi* (shrivel), *ŝtormo* (storm), *ŝtrumpo* (stocking), *ŝvebi* (hover)
trajno (train), *tvisto* (twist)
vjolo (viol), *vringi* (wring)

Syllable-initial cluster only: *fjordo*

Syllable-final clusters only:

abnegacio (abnegation), *obstina* (obstinate)
adjektivo (adjective), *tegmento* (roof), *teĥnika* (technical)
najbaro (neighbour), *rajdi* (ride), *fajfo* (whistle), *bojkoto* (boycott), *sojlo* (threshold), *hejmo* (home), *tiujn* (those, ACC PL), *tajpi* (type), *fajro* (fare), *pejzaĝo* (landscape)
sukceso (success) , *akĉento* (accent), *ekstera* (outside), *ekstreme* (extremely), *akto* (act), *elektra* (electric), *ekzerco* (exercise).
balbuti (stammer), *falĉi* (mow), *faldo* (fold), *kaldrono* (cauldron), *elfo* (elf), *Belgo* (Belgian), *malgraŭ* (despite), *velki* (fade), *filmo* (film), *kelnero* (waiter), *helpo* (help), *falsa* (false), *salti* (jump), *disvolvo* (development), *balzamo* (balm)
ambaŭ (both) *membro* (member), *cirkumcido* (circumcision), *fremda* (foreign), *stamfi* (stamp), *kimro* (Welsh)
alianco (alliance), *vendi* (sell), *infano* (child), *angulo* (angle) *angla* (English), *aranĝi* (arrange), *anheli* (gasp), *sinjoro* (mister), *venko* (victory), *dimensio* (dimension), *instali* (install), *inventi* (invent), *recenzo* (review).
sceptro (sceptre)
varbi (recruit), *komerco* (business), *serĉi* (search), *birdo* (bird), *perfekta* (perfect), *forgesi* (forget), *urĝo* (emergency), *arĥaika* (archaic), *verki* (compose), *perlo* (pearl), *vermo* (worm), *forno* (oven), *harpo* (harp), *diversa* (diverse) *verŝi* (pour), *arto* (art), *konservado* (conservation), *kurzo* (rate of exchange)
lesbo (lesbian), *disĉiplo* (disciple)
matĉo (match), *ritmo* (rhythm), *etna* (ethnic)
bovlo (bowl)

Although Esperanto appears to contain a number of rare consonant clusters, these are often due to one-off forms such as *nh* (*anheli* to gasp), *mr* (*kimra* Welsh), *sĉ* (*disĉiplo* disciple), *tĉ* (*matĉo* match: adopted to distinguish from *maĉo* chewing). Many are also 'false' clusters, created by the conjunction of two syllables of the form VC + CV

(*ad/moni* to admonish, *eks/cepti* to except, *obs/cene* obscenely). A similar but much more widespread formation of long consonant clusters takes place in word compounds and derivations: *sub/ĉiela* under the sky, *sub/donis* contributed / bribed, *karb/polvero* charcoal dust, *etern/brilaj* forever-shining, *inter/trakti* to deal together *gas/bremsis*, *ruĝ/brika* red-brick, *validec/dato* expiry date, *mond/fama* world-famous. It should also be noted that there are in virtually no word-end clusters in Esperanto, excepting the effect of rule 16, which leads to the full range of word-final clusters. In most cases however, word-end clusters from other languages become word-middle clusters followed by a vowel (usually a functional morpheme), as in *sendi* (to send), *arbo* ('arbre', tree), *ŝranko* ('Schrank', a cupboard). There are one or two possible exceptions, such as *ajn* (indeterminate particle). Although Esperanto is often said to sound like Italian, the range of consonant clusters we have set out above and the many compound words we do not have the space to list here give the language a more Slavic quality.

Wells (1989) has pointed out certain common developments of the phonology of European languages which undermine the strict one-symbol / one-sound principle. For example, position affects all sounds, creating allophones as in the /n/ of *bano* 'bath' as opposed to /ŋ/ in *banko* 'bank'. Zamenhof referred to the vowel length and quality of Esperanto vowels as being intermediate between long and short, open and closed (i.e. the mouth for vowels <e> and <o> should be half way between open /ɛ/ and /ɔ/ and closed /e/ and /o/). Zamenhof is said to have pronounced his vowels invariably regardless of position. However, Janton (1994) states that there is almost free variation for these vowels amongst users (he cites /bela/ and /bɛlɑ/) while Kalocsay and Waringhien (1985) state that pronunciation depends on position and that open and closed versions of <e> and <o> are permitted in open and closed syllables only (i.e. /veni/ 'to come' but /vɛndi/ 'to sell', /modo/ 'fashion' but /mɔndo/ 'world'). Wells (1989) argues that this is a useful description, especially convenient for speakers of languages such as French with a range of vowel-widths, but may not be as useful for others such as Japanese or Greek. Most Esperantists presumably speak with a 'local accent' and although little research has been done on national variation, there is some evidence that speakers of Italian impose their own local dialectal traits on the language (Migliorini 1985:67).

The stress pattern of Esperanto is theoretically fixed to an accent on the penultimate syllable in all polysyllabic words (i.e. Rule 10: the accent is placed on the last syllable but one.). Hence *granda* is pronounced /gránda/, usually with a correspondingly slight lengthening of the vowel. This predictable syllable-timed intonation pattern allows for words to be distinguished in speech. Because of the morphology of the language all lexical items are compound words (i.e. even root words should include the word class morphemes *–a, -e, -i, -o*) and therefore carry one stressed accent. For example:

Written form: *Ili parolas la internacian lingvon Esperanto.*

Pronunciation: /íli parólas la internatsían língvon esperánto/

                       (They speak the international language Esperanto).

The intonation pattern typically involves rising intonation for interrogatives (Kalocsay and Waringhien 1985), although to avoid ambiguity polar interrogatives ('yes-no' questions) are explicitly marked by the particle *ĉu,* a usage derived from Polish. It has been noted that at the lexical level, the similarity of certain words to national language forms causes speakers to adopt varying stress patterns. For example, French speakers often emphasize the last syllable, so *internacia* is sometimes pronounced /internatsiá/ instead of /internatsía/ (with a corresponding consonantal /j/ instead of /i/ at the end) whereas English speakers may introduce a pretonic accent and say /internátsia/. In connected speech, where the accented vowel is not stressed strongly, all speakers may also be heard to move the accent leftwards in trisyllabic words as in /áliel/ instead of /alíel/ (*aliel* 'otherwise') or /áŋkoraw/ instead of /aŋkóraw/ (*ankoraŭ* still, yet). The status of these words lies somewhere in between lexical and grammatical items and so there may be a tendency to treat them differently. It has long been noted (Venture 1977), that speakers of various nationalities diphthongize the vowel sequences in correlatives and their derived forms (*kiel* -> */kjel/, ĉiuspeca* /tʃju'spetsa/). Morphological changes in words may also affect pronunciation, and national accents are often cited as causing occasional problems for communication (Wells 1989). In addition, Kalocsay (1963) claimed that the emergent intonation patterns in Esperanto were a natural evolution away from the general principle that Zamenhof had laid down and were following patterns observable in Italian and possibly other languages.

He points out that Zamenhof maintained a strict monotonic pronunciation with half-closed vowels and the penultimate accent in a word such as /libertémpo/ 'free-time, holiday'. One problem Kalocsay identified was that the potential reading of rule 10 is to place an accent on each morphological root rather than just towards the end of a compound word. Thus the official root *liber-* is written *libera* (with the accent on the *e*) and it could be argued that this should be maintained in compounds, thus giving us /libér'témpo/. In current usage however, Kalocsay claimed that speakers introduce a slight pretonic accent instead: /líbertémpo/ (where the penultimate E also becomes a long /ɛ/).

Notwithstanding national differences however, Esperanto's system of intonation is enormously streamlined compared with the ethnic languages and later planned language projects, such as Interlingua. Consistency of morphology and pronunciation is the priority in Esperanto, whereas the other naturalistic systems often aim for maximum recognizability. For example, Migliorini (1985) compares the penultimate syllable rule in Esperanto with Occidental, which introduces new accents in relation to morphological change, and at the same time imposes unpredictable but international norms of pronunciation (giving *anatóm, espíscop, hypérbol, paraból...*).

2 MORPHOLOGY.

2.1 Typology.

Esperanto can classified in a number of ways. In social and historical terms the language belongs to a long tradition of 'planned' or invented languages, that is a language designed for some specific purpose. Esperanto can be considered from a purely lexical viewpoint as a Romance language (a language which is lexically derived from Latin), or from a morphological point of view as an agglutinating language (a language in which word compounds can be created very consistently) Linguistically, Esperanto may be considered to be a Creole (as a contact language and a streamlined version of an original European language).

1) Esperanto as a planned language.

Linguists identify two basic classes of planned language: schematic and naturalistic.

Schematic (also known as *'a priori'*). Schematic systems create a language of images or concepts rather than words. A schematic word is an abstract symbol, and is unlikely to resemble any recognizable language. Schematic grammars are comprehensive but also complex and unwieldy. To be of any use in communication, such a grammar must be bound by certain linguistic universals: for example the syntax can not avoid recursivity and government. Most schematic projects relied on graphic or sound-based systems of symbolic representation, such as Sudre's musical 'Solresol' in 1817. Most schematic languages were developed around the 18th Century (Eco 1997), although some modern versions were later developed (for example, Hans Freudenthal's 'Lingua Cosmica' designed for interacting with aliens).

Naturalistic. (also known as *'a posteriori'*): A naturalistic language project is based on the common features of a series of existing natural languages. Naturalistic words are designed to be maximally recognizable, but depend on the classically educated speaker's prior knowledge of at least one of the natural languages concerned (most schemes are therefore Eurocentric). It is difficult for naturalistic systems to eliminate

idiomatic structures and historical inconsistencies which hinder comprehension between even very closely related languages. Most naturalistic schemes emerged (and died out) in the 19th century, although several schemes based on Esperanto were developed in the first half of the 20th century. Naturalistic systems range from simplified versions of one language (Basic English, Latine sine Flexione) to schemes which standardize across language families, usually the Romance language family (Idiom Neutral, Interlingua).

Esperanto is usually classed as a naturalistic scheme. There are however a small number of schematic features, notably the 'correlatives', the word class morphemes (*-o, -a, -i, -e*) and the reductionist tendency to use certain morphemes to form whole classes of lexical items (as in the use of *mal-* to form opposites).

2) Esperanto as an agglutinating Romance language.

Esperanto is usually considered to be an agglutinating language (such as Hungarian, Turkish or Japanese) with a romance vocabulary (based on Latin). However, different classifications of Esperanto have been advanced, largely because Esperanto's morphological system does not behave in quite the same way as typical agglutinating and Latin-based languages. Here we review a number of ways in which linguists and Esperantists have addressed the issue.

Typological classification is traditionally oriented around the ways in which words are constructed and conceptualized in the language. Words in English, Chinese and many West African languages can be used on their own with no other morphological endings (these languages are said to be 'word-based'), while in languages such as Latin, Russian and Japanese it is impossible to strip away explicit grammatical affixes from most lexical words (these are 'stem-based' languages). There has been much controversy over whether Esperanto is word- or stem-based, with many linguists claiming that Esperanto is essentially stem based. This is partly because verbal and other lexical endings are obligatory in Esperanto and also because the interpretation of lexical roots is limited by their 'natural' word class (Saussure 1910, Kalocsay 1931, Schubert 1989b, 1993). Others, notably Piron (1991) and Janton (1994) have argued that Esperanto's 'word roots' have no inherent word class, placing greater emphasis on

the function and freedom of use of the 'word-class' morphemes (nominal -*o*, adjectival -*a,* verbal -*i*). Notwithstanding this debate, as with most languages Esperanto has features which are typical of all the traditional classifications (flexional, agglutinating, isolating). Esperanto can be further categorized in terms of two dimensions: the lexical-syntactic distinction <u>analytic - synthetic</u> (the extent to which ideas are expressed by a combination of existing words or by resorting to new word forms) and the morphological distinction <u>agglutinating – fusional</u> (the extent to which morphemes have one or more functions in the language). Using Greenberg's basic measure of this distinction for other languages, Wells (1989) suggests that Esperanto stands around half way on the first scale with some synthetic features, but on the second scale is very highly agglutinating.

Wells (1989) and Cherpillod (1989) set out the case for considering Esperanto to be an agglutinating language because of its system of monofunctional morphemes. The argument that Esperanto is like Kiswahili or Turkish is based on the fact that Esperanto's morphemes are used consistently for the same grammatical features, a difference that distinguishes Esperanto markedly from the Romance language family. For example, the fused nature of many Spanish morphemes means that no one category has a distinctive form, so while –*o* is usually associated with the first person in present tense verbs (*canto*: I sing), it can not be used to identify verbs consistently because –*o* is also a signal of masculine nouns or adjectives. In addition, in the past imperfect form *yo cantaba* 'I was singing', the form -*aba* is a fused morpheme which signals at the same time 'singular, first person' and 'past tense, imperfect aspect', while the pronoun *yo* signals first person and is optional. In Esperanto the equivalent form is *mi kantadis*, but all three morphemes are consistent in Esperanto: they remain the same for all words and in combination with all grammatical categories. It is for this reason that the four forms *mi / kant / ad / is* are all known as roots (primitive words) regardless of grammatical or lexical function (they are also known in Esperanto as 'official roots' i.e. they were part of the original design of the language). Another difference with Spanish is that the pronoun *mi* and past tense marker -*is* are obligatory and only the aspect morpheme –*ad*- is grammatically optional. This degree of morphological signaling for tense and person establishes Esperanto as partly analytical (the pronoun stands free of the verb and its position is largely free) and partly synthetic (the tense marker is obligatory and fixed in position). In addition, the

obligatory signaling of grammatical word class for lexical items (nominal, adjectival, verbal) in Esperanto is more akin to some Amerindian and African languages where classifiers and other particles are used for this function. It might even be argued that the 'morpheme effect' (described below in section 2.4) may be a morphological feature that Esperanto shares with a incorporating languages (including Polynesian or Amerindian languages). As in Esperanto, many of these languages signal word class explicitly, except when a root word is used in compounds (to give one example from Chukchee: /nə teŋ gin/, 'ADJ +good +ADJ' which becomes /tə teŋ ŋelgə l?u k/ 'I +good +hide +saw +I', where the adjective root loses its ADJ morphemes).

Although Esperanto's morphology appears to be highly agglutinating, Piron (1991) considers several non-agglutinating aspects of the language, including the freedom of word derivation and monomorphism that are reminiscent of isolating languages such as Chinese or Vietnamese. Piron's thesis is that Esperanto only resembles agglutinating languages superficially, and that there are a number of important differences, not least of which the fact that agglutinating languages impose sound changes such as vocal harmony and consonant weakening. Piron points out that each lexical and functional morpheme in Esperanto has a fixed form (in principle there are no 'morphs' or variants) and that each has a relatively fixed semantic meaning with a variable word class. The word in an agglutinating language is stem-based and can not be used without obligatory grammatical endings, whereas Esperanto may be seen as word-based with many independent word forms which cross the normal boundaries of lexical word formation in agglutinating languages and other stem-based languages (for example, the use of adjectives as predicative verbs, the potential to omit word class morphemes such as –o, the fundamental and free role of word class morphemes – termed 'functional morphemes' below). More fundamentally, Chinese, English and Esperanto are said to share relatively fixed word order in comparison with stem-based languages such as Latin or Russian. This is especially true in terms of the ordering of preposition, noun, verb and adverbial phrases, all consequences of an isolating structure which downplays the role of morphology in assigning grammatical functions. In Esperanto as in Chinese, the verb lacks conjugation for person and number, another isolating property. Similarly, other features such as prepositions and grammatical items may be used to form new words. Unlike Chinese and other

isolating systems however, Esperanto has obligatory grammatical features such as the obligatory inclusion of tense, transitivity, and reflexive clitics and so on in the verb phrase, as Piron himself points out (1991).

Finally, Esperanto has inherited a small but influential number of morphological features from inflecting or fusional languages. Kalocsay (1931) was one of the first linguists to point out the numerous influences of the Slavic languages on the application of Esperanto's morpho-syntactic system (including the role of prepositional phrases in noun groups, transitivity, verbal morphology and participle usage). The main synthetic features in Esperanto can be seen in the agreement between adjectives and nouns for case and number (involving the –a, -o forms: *estas diversaj problemoj* 'there are several problems') as well as demonstratives in –iu (*ili komprenas tiujn problemojn* 'they understand those problems'). Esperanto can not be considered fusional or synthetic however because in general a morpheme in Esperanto never changes the form of the word root. There is one exception to this: the morphemes –nj- and -ĉj- are used to convert proper nouns into familiar names and thus change the word root into a truncated stem (*Vilhelmo* William -> *Vilĉjo* Bill, *Aneta* 'Annette' -> *Anjo* Annie). The borrowing of words from various languages also leads to some trace features of 'fusion' in the vocabulary. For example, different morphs are used for the same concept in *direkti* 'to direct', *direktoro* 'director', *direkcio* 'directorate'. These words can also be formed by Esperanto's general rules of word formation, although the meanings of Esperanto's derived (analytical) forms are less restricted (*direkti – direktisto – direktistaro* Piron 1991:12). Esperanto therefore has a dual lexicon: borrowed international words which are often synthetic, and regularly formed derivations, which are less recognizable but display consistent morphological analyticity. We examine the nature of word-compound in Esperanto in further sections below.

3) Esperanto as a Creole.

Esperanto can be described as an artificial Creole, lexically derived from a combination of the main Western European languages but with morpho-syntactic structures that have been systematically generalized. This is a theoretical viewpoint, however, and Esperanto has little in common with true Creoles, which are more

accurately defined by their social and historical circumstances: namely they are developments of the pidgins spoken by slaves in former European colonies (such as the French Caribbean Creoles in Haiti and Martinique). Pidgins as such are usually developed to cope with certain types of interaction, such as exchange and negotiation, but once adopted by a community these systems are elaborated as Creoles. More realisitcally, Esperanto can be compared to languages which have been revied and extended for social or political reasons, such as Ivrit (a form of classical Hebrew adopted in modern Israel: Versteegh 1993). Adopted or revived languages often undergo a process of 'creolization' which may be compared with the development of grammatical innovations in Esperanto. The relationship between Esperanto and Creoles is therefore more abstract: the transformation of Pidgins into Creoles may be reflected in the development of Esperanto from a schematic design to a relatively widely-used language. The use of Esperanto in families with small children is particularly reminiscent of the adoption of Pidgins as a language of the family, as well as bilingual situations elsewhere (Versteegh 1993, Rasič 1994).

From the point of view of language features, Esperanto shares some but not all the typical characteristics of Creole languages. Here we set out the main features of Romance-based Creoles (based on Green 1988 and Versteegh 1991):

Similarities with Creoles.

1- Creoles have no **inflectional morphology**. Creoles typically lose the gender, number and case systems of their parent languages. Esperanto similarly has few inflections, although the grammatical morphology of the language is much more complex than that of many Creoles (for example word class endings for nominals, adjectives and verbs). The plural and accusative endings of Esperanto (*j* and *n*) are uncommon in Creoles and they involve obligatory agreement, which is also not a feature of Creole languages.

2- Creoles have consistent **verbal markers** for tense-mood-aspect (TMA), most often preceding the verb: e.g. 'k̲a̲ parti' (PAST + go) 'went' = *foriri̲s̲* (go + PAST). Esperanto's TMA markers resemble the simplicity of Creoles in that they do not agree with the subject, and may be used interchangeably (in some unofficial forms aspect + tense can be combined: see section 3.4). Creoles and Esperanto also make a clear

distinction between punctual / habitual aspect (optionally marked in Esperanto by endings such as the habitual *–ad-* or the progressive passive participle *–ata-*). In this respect Esperanto and Creoles have greatly simplified the verbal systems of languages such as French or German. Creoles differ however in that they often do not require a present tense marker, and often have no a future tense marker. Creoles also have no equivalent of the imperative / subjunctive *–u* in Esperanto.

3- Creoles have **agglutinated** nominals, usually involving analytical word compounds such as 'Tok Pisin' (literally 'pidgin talk' or language). This is often the case in basic ('basilectal') Creoles and Pidgins, although it tends to not to be as widespread as in Esperanto (see section 2.2, below).

4- Creoles use **intensifiers** and ideophones as productive parts of the vocabulary. This is also common in some European languages (diminutives such as '-je' in Dutch or augmentatives such as '-issimo' in Italian). English uses as redundant forms and fixed expressions to express intensity: '(highly +) strung, full (+ to the brim), (high) wind'. For this function, Esperanto uses regular suffixes such as *–em-* and *–eg-* (*ekscitema*, *plenega*, *ventego*) and similarly the diminutive *-et-*: *libro / libreto* book / booklet, *lago / lageto* lake / pond, *kolera / kolereta* angry / miffed or annoyed.

Differences with Creoles.

1- Creoles have little productive **derivational morphology**. Creoles rely on borrowings or on heavy use of analytical paraphrases, whereas Esperanto differs radically in this respect: the derivational morphology of Esperanto is one of the most productive areas of the grammar. In many Creoles, the grammatical function of words must be gathered from context and is not signaled morphologically. For example, in Bichelamar, an English-based Creole, *mimi* is used for I, me, my and mine, whereas Esperanto has *mi, min, mia,* with additional endings for case and number *mian, miaj, miajn*.

2 Creoles typically signal the role of complements in a sentence by using clitics, obligatory pronouns signaling grammatical function. The function is similar in French left-dislocations ('Lui, c'est mon patron' 'Him, that's my boss'), but this feature is completely absent from Esperanto.

3- Creoles typically do not have **copula** verbs in present / generic contexts. In contrast, Esperanto has copular verbs for attributive clauses (*havi* to have, *esti* to be)

although zero-copular predicative constructions can be regularly used as in *mia brako malvarmas* my arm is cold (= literally 'my arm cold + PRES').

4- Creoles have **serial verbs** (for example, the French-based 'pren aller' = 'to bring'). Verbal endings in Esperanto such as *-ig* (causative) may be considered to have a similar analytical function. However, this is not a common feature of Esperanto (except for projecting verbs: *iris vidi* went to see, *igis lin fari* made him do…).

5- Creoles use **reduplication**, especially for emphasis or to signal the plural (for example 'emi-emi' is the plural of breast in Malay, 'bay and bay' is a future marker in many English Creoles). The European languages and Esperanto use this sparingly, usually in cases of emphasis or for specific idiomatic usages (there are some examples however: *finfine* at long last from vs. *fine* at last, *plenplena* 'full to the brim', from *plena* full, but a regular form *plenega* 'very full' also exists).

While we have focused on simple linguistic features, we would claim that Esperanto shares certain more fundamental features with Creoles. Indeed many languages share Creole-like characteristics, since all languages are subject to contact, invasion, migration and other factors that are likely to lead to simplification and rapid grammatical change. For example, the development of a streamlined morphology in modern English with its can be considered to be in part a consequence of the contact and convergence between the related Anglo-Saxon and Scandinavian dialects, as well as the presence of Norman French in the development of the language. This does not mean that English or Esperanto are 'simple'; a loaded term implying linguistic poverty and lack of sophistication (and a term often abused by propagandists). Hagège (1985) makes the point that the grammatical streamlining of languages such as English or Esperanto, as well as Creoles, is more a stylistic feature than a typological one. Instead, Hagège invokes the principle of <u>economy</u>. Creole languages, informal spoken varieties of ethnic languages, and planned languages such as Esperanto all tend to reduce systematically the range of paraphrases available. For example, standard French has three question forms (Tu viens? Viens-tu? Est-ce que tu viens?), while colloquial French prefers the same form as the indicative but with rising intonation: 'Tu viens?'. Esperanto similarly uses the indicative order with an obligatory particle: <u>*Ĉu vi venas?*</u> with only optional rising intonation. Similarly, French vernacular is currently rapidly simplifying minor verb forms (verbs in *-ir, -oir* and *–re*) an replacing them by the general ending *–er*: voir -> visionner (*vidi*, to see, to view), finir ->

finaliser (*fini* finish, finalise), parfaire -> perfectionner (*perfektigi* to perfect, make perfect), recevoir -> réceptionner (*ricevi* to receive, to take receipt of), sélectionner -> choisir (*elekti* to choose, select). Rather than making the language poorer however, the effect in French is to form informal words which also have a more specific meaning and which can be contrasted with the original form. It is noticeable that for all these examples English uses paraphrases or Latinate words with a similar stylistic effect to French (shifting from an unmarked to a marked form). This form of lexical dispersion is also present in Esperanto using a different process: the formation of analytical compounds (*montri* to show vs. *vidigi* to show, to cause to see, *akiri* acquire vs. *havigi* to get, to cause to have).

While it is true that tense and aspect markers, grammatical gender and the range of cases are typically dramatically simplified in Creoles, it is also true that these languages develop their own areas of grammatical complexity. Most French-based Creoles eliminate grammatical gender, for example but introduce obligatory transitivity markers or establish deictic systems (of proximity and reference, sometimes involving dual as well as plural number) that are very different to the original colonial language. Esperanto similarly imposes consistent but not necessarily 'simple' systems, such as causative / ergative markers on verbs (*-ig / -iĝ*), adjectival agreement for case and number, the opposition zero-article / article (*la*), obligatory marking of attributive adjectives (*-a*) and zero-referent modifiers (*-e*), use of a 'subjunctive' form *-u* after certain conjunctive expressions etc. These systems are still 'economical' in the sense that they are applied across the language without exception, but they are nonetheless difficult for learners who do not recognize these distinctions in their own language.

As well as economy, Pidgins and Creoles display a high degree of <u>analyticity</u>:, whereby a complex idea will be broken down and expressed by a series of more simple concepts. For example, 'gathered something for me' is expressed (in Haitian) as 'went come-back pick give me'. Vernacular languages and Esperanto do share similar processes to this, although not to the same degree. English has a complex vocabulary for example, which is divided into high and low styles: i.e. Latinate forms 'to ascend', 'to transfer', 'violence' are 'High' and can be contrasted with 'Low' Anglo-Saxon forms which are relatively analytical: 'to go up, to hand over, show of strength' (the 'low'

forms are not necessarily more informal, but they are less likely to be used in 'high' contexts, such as legal texts: we consider them therefore to be 'marked' rather than just 'informal'). Esperanto similarly prefers to express these concepts by analytical derivations which are similar to the English low forms: *supreniri* (to go up), *enmanigi* (to place in the hand), *perforto* (by means of strength). It is noticeable that possible 'High' forms also exist in Esperanto: *ascendi*, *transferi / transdoni,* but these are very infrequently used (in fact these Latin-based forms are considered to be neologisms relative to the compound 'Esperanto' forms). In many cases in Esperanto there is no Latin or 'synthetic' equivalent (*perforto* is the only commonly used word for 'violence'). To take another example, vernacular French has no equivalent for the Anglo-Saxon 'Low' words in English, but analytical expressions are common in colloquial contexts, and these also happen to be reflected in Esperanto, usually in derived or compound words: partir -> s'en aller / *foriri* (to go away), tournure -> façon de parler / *parolmaniero* (a speech-style, expression), ménagère -> bonne à tout faire (literally 'woman to do everything') / *servistino* (house-maid, 'serve + worker + feminine'). Esperanto can therefore be said to be relatively more analytical than some other languages, although this classification again centers on style.

Another Creole-like characteristic shared by Esperanto and vernacular languages is motivation, the tendency to cluster expressions around a basic term. For example, standard English and French use unrelated (non-motivated) terms for good / bad = bon / mauvais, whereas Esperanto and certain pidgins use words derived from one central term (here compared with Neo-Melanesian): gut / nogut = *bona / malbona*. Few languages choose to use the negative word as the basic term, but there are exceptions. In informal French and English, it is common to express these ideas analytically and to use the negative word as the basic comparator (not bad = pas mal), as well as to prefer analytical versions of comparisions (same / not the same = same / different, in French: pire / plus mal = worse / more bad). Esperanto is more pidgin-like in this respect, because the motivated forms are obligatory and not just a question of style (indeed, there is no word for 'different' in this context: *malsama* = 'unsame', *malsimila* = 'dis-similar'). The *mal-* morpheme is very productive in Esperanto, although as we mentioned above other motivated groups of words are also evident in the language (for example, consistent feminine words are formed on the basis of

masculine forms: *viro / virino* man / woman, *dio / diino* God / Goddess, similarly for collectives: *arbo / arbaro* tree / wood, *haro / hararo* hair / head of hair).

Despite major differences in historical context and in basic linguistic features, we have seen that Esperanto shares some abstract and fundamental similarities with Creole languages, and that this is akin to the development of vernacular style in other ethnic and national languages. Nevertheless, it would be wrong to consider Esperanto as a Creole itself.

2.2. Lexical Morphology.

The specific nature of Esperanto's morphology leads us to posit three main classes of word: lexical, functional and grammatical. Esperanto has the following general lexicon, the distinction between 'word classes' in Esperanto being the extent to which the different word roots may combine with free or bound morphemes and each other:

Free items:	Bound morphemes:
Lexical items (all lexical roots)	Lexical morphemes (all derivational morphemes)
Functional Items (prepositions, pronouns, grammatical adverbs, numerals)	Functional morphemes (word class morphemes *-a, -e, -i, -o, -u* e)
Grammatical Items (conjunctions, particles)	Grammatical morphemes (*-n, -j*)

Lexical items. All lexical items are assigned one of three grammatical word classes and combine with a bound morpheme: <u>labor-</u>o ('work' nominal), <u>rapid-</u>a ('quick' adjective), <u>manĝ-</u>i ('to eat' verbal). These lexical roots can change word class freely in conjunction with other bound morphemes (*rapidi*, to be quick, *manĝo* a meal) and can often act as free morphemes (*rapidmanĝi*: 'to eat quickly' or 'to gobble food up'). Most words in typical use involve two morphemes: the root plus a functional morpheme (see derivational morphology below), although monomorphemic lexical items do occur in literary production (the final *-o* is omitted according to rule 16 as in Auld's *'l'animo estas <u>pens'</u> kaj emocio...'* the soul is thought and emotion). While there is no upper limit to the combination of morphemes, four morpheme words represent the

typical maximum in our text corpus (*poŝtoficejo = poŝt- ofic- -ej -o* = post + office + place + noun).

Functional items. This class includes words with an intermediary role between grammatical and lexical items: pronouns (*mi, vi ĉiu...* I, you, every one…), numerals (*unu, du, tri* one, two, three), grammatical adverbs (*tre* very, *baldaŭ* soon, *nur* only, *kiel* as, *tiam* then etc.) and prepositions (*al* to, *trans* across, *pro* because of). These items are considered intermediate because while they are often used in isolation with no modifying elements (as in *ĝi estas kiel tiu*: it is like that one), they can also combine in limited cases with a number of functional morphemes to become full lexical words (*via* yours*, la unua* the first, *la tiama prezidento* the president at the time, *transa* transitional). Prepositions are similarly widely used as prefixes (*aliri*: to go towards, *kunlabori*: to work together).

Grammatical items. In isolation these items have specific, abstract grammatical functions: *la* 'the' determiner, *ke* 'that' conjunction, *ju pli...des pli* 'the more' comparative conjunction, *mem* 'self' emphatic determiner, *eĉ* 'even' emphatic determiner. This category also includes so-called particles such as *ĉu* (interrogative), *ĉi* (deictic), *nu* (interjection). The use of any of these items as morphemes is usually highly restricted (there are one or two exceptions: *ĉi* can at times be fused as demonstrative particle + root, especially adverbs: *ĉi-jare* this year, *ĉi-foje*: this time, and *ĉu* has been attested as a verb root in *ĉuas* 'dither': Piron 1989a:132).

Lexical morphemes. These words function in the same way as lexical items, the only difference being that they are primarily used as affixes in derivational morphology and represent very broad categories. In the dictionary these do not have the same word class assignments as lexical items although their use tends to fall in one or another category (*-in-* 'feminine' nominal, *-et-* 'small' adjectival, *-ad-* 'continuing process' verbal). Lexical morphemes create new lexical items (*malrapida* slow, *pafilo* rifle) or change the scope of the word's meaning (*rapideta* somewhat slow, *mispafo* a misfire). The verbal participle endings are also lexical morphemes (*-ant-, -at-* etc.) in that they affect the semantics of the verb in similar ways to other derivational morphemes and may be used with nominal and other functional morpheme endings. The correlative prefixes (*ki-, ti-, ĉi-, i-, neni-*) can also be seen to be lexical

morphemes which tend to only combine with functional correlative morphemes (although they may have spread to other lexical uses: *neniulo* a nobody, *aliulo* someone else).

Functional morphemes. These include the word class morphemes: nominal *–o*, adjectival *–a*, adverbial *–e*. These elements determine the grammatical function of the root word in any given utterance and in some cases may be 'inflected' by grammatical morphemes (for example, adverbs may take the accusative). According to this definition the correlative suffixes (*-om, -el, -am*) are also functional morphemes (signaling adverbs) and applied only to correlative prefixes. In more restricted contexts, the forms *-u* and *-i* are similarly non-productive functional morphemes which signal pronouns (for correlative prefixes and personal pronouns only) while the correlative ending *-es* signals a possessive or relative pronoun (see 'correlatives', below). The verbal endings also count as functional morphemes although they do not allow for further grammatical morphology or suffixes: *-i* (infinitive)*, -u* (imperative, volitive), *-as* (present tense), *-is* (past tense), *-os* (future tense), *-us* (conditional) Their internal consistency, in particular the consistent use of the form *–s* (= 'finite verb') could be considered a form of bound grammatical morphology. The ending *-aŭ* also deserves some mention here as a neutral word class morpheme which has survived from Zamenhof's earlier project designs. It is bound and non-productive, although it appears in a number of high frequency functional words (prepositions: *antaŭ* 'before', specifying adverbs *ankaŭ* 'also').

Grammatical morphemes. This involves two 'bound' inflections: plural number *–j* and accusative case *–n*. They are involved with grammatical agreement and have no effect on the grammatical word class of the word they are bound with. They can only be applied to functional morphemes and no further morphemes can follow them (although *–n* may find its way into compounds such as *antaŭeniri*: to go forwards).

2.3 Grammatical morphology.

In this section we examine in closer detail the applications of the Sixteen Rules on the grammatical morphology of Esperanto. Many of the observations below involve the relationship between morphology and syntax, and some features (notably pronouns and particles) are discussed later, under syntax. When discussing appropriateness or grammatical acceptability it helps to have in mind a general norm that is valid across a large sample of language. All examples from this point on are therefore taken directly from the texts in the corpus (see Appendix). For some features we set out the statistical distribution of different linguistic forms in Esperanto. Concordances and lists of word frequencies made available by computer can be compared with early concordances of Zamenhof's writings (in particular the *Fundamento*: Wackrill 1907) and can demonstrate patterns of usage that are not accessible by introspection or manual searching alone. Some of the most interesting questions about Esperanto involve the extent to which the linguistic system is exploited in the same way as other languages. The following questions are assumed: is the language really as consistent as is claimed by its supporters? Are some parts of Zamenhof's designs underused? To what extent have exceptions or innovations (such as the innovative verb forms *–intas, -atas*) become accepted by the community as a whole and in what way do they impinge upon the rest of the system?

1) **Nouns**. (*Fundamento* Rule 2).

The morphological mark of all nouns is *–o,* plural *–oj.* This suffix may in principle be applied to all lexical word classes (thus *ŝuldi* to owe, *ŝuldo* a debt). It can be applied to some functional items - usually prepositions (*antaŭ-* before *antaŭo*: front) or correlative prefixes (*tio* 'that'). There are two grammatical cases: nominative (no change) and accusative (using *–n*). Thus: *jen radikala decido* here's a radical decision, *transformi la demokratiajn principojn* to transform democratic principles. The accusative is also used in conjunction with or as a replacement for prepositions in three main areas of application:

1) direction (*paŝis trans la fajreto<u>n</u>*: stepped straight across the fire)
2) duration of time (*atendu unu horo<u>n</u>* wait (for) one hour)
3) measurement (*pezas 200 kilogramoj<u>n</u>* weighs 200 kilograms).

Other roles are expressed by a combination of preposition and unmarked case: *vi povas sterni ĝin <u>en</u> la mezo*: 'you can lay it out in the middle'.

Some exceptions emerged for Rule 2. An apparent exception involves the use of adjectives to name languages (*la franca, la itala* – French, Italian) where the noun *lingvo* is conventionally omitted. Similar omissions occur with dates (*la duan [tagon] de decembro*: the second [day] of December). In addition, feminine proper nouns such as 'Johana' were initially expressed using the feminine morpheme *–in-* i.e. *Johanino*, however a nominal ending *-a* was later preferred under the influence of the European languages. In current usage all proper nouns, especially uncommon ones ending in consonants are regularly left in the original pronunciation or transcription, despite attempts to Esperantize foreign spellings (a tendency known as *'ŝulcismo'*).

2) Adjectives. (*Fundamento* Rule 3).

The morphological mark of adjectives is *–a,* plural *–aj.* Adjectives agree for plural number and accusative case (*li ĝuis tiu<u>jn</u> longe admirataj<u>n</u> lipoj<u>n</u>* 'he enjoyed those long-admired lips'). This functional morpheme may be applied to all open-word classes (thus *vento* wind, *venta* windy). It can also be used with prepositions and other functional items (*nur* only, *nura* mere, *post* after, *posta* later, latter), and with pronouns *-a* forms possessive pronouns and determiners: *mi* me, *mia* my, mine. Comparatives are formed with *pli*, or superlative *plej*, the standard for comparison is signaled by the conjunction *ol* and placed at the end of the phrase (*ĝi estas <u>pli</u> granda <u>ol</u> tio* 'it is bigger than that'). *La plej* is the superlative for adjectives, but *plej* alone specifies adverbs. It is noticeable that in the standard expression *kiel eble plej rapide* 'as fast as possible' the standard comparator is an adverb and placed before the variable, a word order ascribed to Slavic languages.

In Esperanto the role of adjectives is restricted in two respects. In the first instance they are nominally bound (i.e. to *o*-morpheme, pronouns and the like), that is to say

they can not be attributes of empty subjects or verbs and clauses serving as subjects. In these cases the *e*-form (adverb) is used instead: *estas vintre* 'it is wintery', *plendi estas vane* 'to complain is futile' (for more examples see syntax, below). This has been seen as proof of the independent nature of lexical roots in Esperanto (the morphemes *e* or *a* do not change the basic adjectival meaning of the original word) and has resulted in hesitation in cases where the adjective can either modify a subject or complement (as in *ŝi aspektas hezitema*: she looks hesitant) or where the adjective is modifying the clause in which case it should be realized as an adverb, as in *tio aspektus strange* 'that would appear (to be) strange'.

Secondly, there is a tendency for Esperantists (following the early writer Grabowski) to replace predicative adjectives by active verbs. Kalocsay (1963:56) cites Grabowksi's *prudenti* (instead of *esti prudenta* 'to be prudent') and *preti* (instead of *esti preta* 'to be ready'). In the corpus the distribution of predicative verbs is large but dependent on a limited number of high frequency verbs. In a sample of 1400 present tense verbs 203 were found to be derived from adjectives. The following list gives a ratio representing the verbal use of some frequent predicative verbs (predicative vs. adjectival use: the second figure only includes predicative attributes of 'esti'):

abundas (is abundant 6:2)	*certas* (is certain 27:20)	*fidelas* (is faithful 22:2)
fieras (is proud 7:0)	*gravas* (is important 67:35)	*kapablas* (is able 54:4)
necesas (is necessary 198:13)	*pretas* (is ready 52:32)	*sufiĉas* (is sufficient 46:1).

The semantics of these verbs appears to suggest a general tendency to use adjectival predicates for abstract states of potential, necessity or quantity (as opposed to more canonical adjectives of physical size, beauty, color etc.). This suggests that verbs are often preferred for the expression of abstract states of potential and quantity in the language although this innovation affects a restricted number of verbs with similar semantic properties. A more canonical adjective, such as *granda* (large) is used 37 times as predicative attribute (for example: *liaj misŝancoj estis grandaj* his bad luck was huge, *nia mondo estas granda* our world is large) and only three times as an active verb form: *se via malkontento tro grandos* if your unhappiness becomes too much, *kiuj same grandas je 2,44 megabajtoj* which are similarly 2.44 megabytes big, *ĉar ĝi tro grandas* because it's too big. This verb form is infrequent although not

unknown with adjectives such as *stulta* (stupid), *bela* (beautiful) and *bona* (good). This suggests that the verbal use of an adjective root is phraseological: i.e. limited to a consistent set of lexical items.

3) Verbs (*Fundamento* Rule 6).

Active verbs signal tense (time from the point of view of the clause) and participle forms signal aspect (duration of action). Verbs do not change for person or number. The morphological mark of infinitive verbs is –*i*. The mark of the imperative (or 'volitive' as it is known by some Esperantists) is –*u*. The four active forms of verbs are bound morphemes: -*as, -is, -os, -us*. There are three participle morphemes (-*ant* progressive, -*int*- perfective, -*ont* prospective) and three passive forms (-*at*- progressive, -*it*- perfective, -*ot*- prospective). These combine with other words as free morphemes. Participles are typically expressed as adjectives in conjunction with the verb *esti* (to be). We demonstrate the basic use of these morphemes below for the verb *manĝi* (to eat):

Active forms: *li manĝas* (he eats, is eating: present), *ŝi manĝis* (she ate, has eaten: past), *ili manĝos* (they will eat: future), *vi manĝus* (you would eat: conditional).

Active participle forms: *ŝi estas manĝanta* (she is eating: progressive), *mi estas manĝinta* (I have eaten: perfective), *vi estas manĝonta* (you are about to eat: prospective).

Passive participle forms: *ĝi estas manĝata* (it is being eaten: passive progressive), *ĝi estas manĝita* (it has been eaten: passive perfective), *ĝi estas manĝota* (it is about to be eaten: passive prospective).

The basic distribution of tenses in the corpus (set out in Appendix 3) shows that there is a skewed distribution between the frequently used narrative tenses (present / past) and the more marked non-actual tenses (future / conditional). The 'volitive' form -*u* is used for traditional imperatives (rare in the written corpus but evident with expressions in prose such as *permesu* 'permit me to…'), and more frequently in constructions which are often analogous to the subjunctive mood in French and other languages. Since the volitive is used in independent as well as subordinate clauses the

term subjunctive is usually avoided. In other languages these expressions often correspond to modality or the infinitive. Since most examples of this involve the verb *esti* 'to be', the following are typical (those constructions which normally require the volitive form are underlined):

1) *Tial mi* **estu** *konciza…*	So let me be brief.
2) *La konkursaĵoj* **estu** *subskribitaj per pseŭdonimoj.*	Competition entries must be signed as pseudonyms.
3) *…absolute necesa <u>por ke</u> la milita ekvilibro ne* **estu** *rompita.*	…absolutely necessary so that the military balance not be disrupted.
4) *Ni <u>deziras, ke</u> la jaro 1996 ne nur* **estu** *la paca jaro...*	We wish for 1996 not only to be the year of peace…
5) *…tial <u>gravas, ke</u> ĝi* **estu** *senescepte respektata*	So it is important that it continues to be respected without exception

[in 5) note the potential untensed English 'subjunctive': *…important that it continu**e** to be…*].

In the corpus, the most frequent use of –*u* is in independent clauses as in examples 1 and 2, especially after adverbs: *tial* (for this reason) and in the expression *ni –u* (let us…). The second most frequent use of the *u*-form is consistently after a small set of subordinating constructions, the most frequent being *por ke* so that, and *tiel ke* such that. Less frequently, a limited group of verbs (as in examples 4 and 5) require the *-u* form in complement clauses (the other examples in the corpus being *ordonas ke* orders that, *penas ke* tries that, *volas ke* wants that, *zorgas ke* is careful that, *konsentas ke* agrees that).

The flexibility of verb forms in Esperanto is a complex area of grammar. As complex verbs, participle forms are used in combination with *esti* (to be). They are typically expressed as adjectives (*-anta, -ata* etc.) but may be used as nouns or adverbs. Participles when expressed as nouns are conventionally taken to refer to persons or animates (*manĝanto* someone eating, *la parolanto* the person speaking, the speaker). Participles behave in the same way as adjectives, with agreement in number and case: *li vidis la viron portant<u>an</u> jakon*: he saw the man (who was) wearing a jacket. Other tenses of *esti* are used with participles to allow for a range of temporal expressions. In addition, complex participle forms can be formed without the use of *esti* (again as with predicative verbs, an innovation attributed to Grabowski). The participle morpheme is thus treated as a derivational suffix, and the form *ĝi estas manĝ<u>ita</u>* (it is eaten) can be reformulated *ĝi manĝ<u>itas.</u>* The active *ŝi estas manĝ<u>inta</u>* (she has eaten)

would be *ŝi manĝintas.* The difference in meaning is therefore not to play down the stative meaning of the participle form but to emphasize the fact that it comes into being as an event. The corpus includes a number of examples:

1 *Tiam, kiam mi vid<u>intis</u> papilion...*	Then, when I had seen a butterfly…
2 *Sed ankaŭ bezon<u>atas</u> persista, diversmaniera laboro...*	But also persistent, varied work is currently needed….
3 *La modernuloj konstru<u>antos</u> sian novan fortreson...*	The modernists will be building their new fortress…
4 *...la ŝipo, kiu nin transport<u>ontis</u> Koreien!*	The ship which was set to transport us to Korea!
5 *Laŭplane la tekstoj kant<u>otis</u> de profesiaj korusoj ...*	According to the plan the texts were about to be sung by professional choruses....

In the corpus as a whole, *esti* + participle is by far the most widespread form (in particular the passive forms), although the compound forms are still relatively frequent:

Compound participle formation:

-ant-	6	-at-	9
-int-	19	-it-	4
-ont-	5	-ot-	1

Esti + **Participle verbal formation:**

-ant-	21	-at-	156
-int-	60	-it-	262
-ont-	15	-ot-	2

The passive participles *–ata / -ita / -ota* mirror the tense forms *–as / -is / -os*. This means that the active form *oni malfermis la pordon* (someone opened the door) can be paraphrased by three passives:

*La pordo estis malferm**ata***	('The door was being opened' – concurrent, progressive event).
*La pordo estis malferm**ita***	('The door was opened' – punctual, perfective event).
*La pordo estis malferm**ota***	('The door was about to be opened' – predicted, prospective event)

When the corpus is consulted, we find that   *-ata* is used in around one third of all cases (plural and accusative forms have been included):

-ita / 5527   *-ata* / 2060   *-ota* / 61.

The progressive form *–ata* is a usage derived from the Slavic languages but does not resemble usage in most western European languages. During the 1960s linguists began to debate the interpretation of participles in what is known as the *ata-ita* debate.

Some Esperantists were concerned that the participle morphemes allowed for differing interpretations, as in the example of *okupita* (occupied), which could be seen as having either the aspectual, state-based meaning 'occupied, and now deserted' or the sequential event-based meaning 'occupied, and yet to be vacated'. However, it is fairly clear from this example that in general *okupita* is likely to be used to describe the event of occupation, regardless of the aspectual connotations of *–-ita* as 'completed, finished'. As Kalocsay and Waringhien noted (1985: 152), the semantics of the verb determine which interpretation is to be preferred in conjunction with the aspect morpheme and the morpheme does not exert one single interpretation. The implication is that Esperanto classfies verb semantics as either events or states, and that both types of verb have a standard interpretation with each participle. Thus for verbs of 'result' (or 'events') the unmarked form is *ita*: *konstruita* built, *farita* made, *perdita* lost (*-ata* would specifically draw attention to progressive aspect) while for verbs of 'duration' (or 'states') the unmarked usual form is *–ata*: *konata* known, *memorata* remembered, *tenata* held (*-ita* would be used to draw attention to the fact that these were finished states). This is similar to the situation in Russian, where imperfective verbs are seen as verbs of duration and normally take the 'progressive passive' equivalent of *–ata*.

The following examples from the corpus show typical usage:

-ata-

1 *...ŝi estas kon<u>ata</u> pro fikcio...*	She is (still) known for fiction…
2 *…Respondo estas garanti<u>ata</u>.*	A reply is (currently) guaranteed.
3 *Jene funkciis la sistemo uz<u>ata</u>*	This was the system in use when I was there.
kiam mi estis tie.	
4 *Sed la insulo montriĝis neloĝ<u>ata</u>.*	But the island showed itself to be (currently) uninhabited.
5 *Tiu lasta estis konsider<u>ata</u> de la reganta.*	The latter was being considered by the
superularo	governing supreme body

It can be seen that certain verbs are almost exclusively used in this form in the corpus (two of which, *bezonata* needed, *konata* (well) known, make up one fifth of all the occurrences of *-ata*) and overall there is a preponderance of verbs denoting debate or proposal (*konsiderata* considered, *proponata* proposed, *garantiata* guaranteed) as well as a number of related verbs prefixed by *pri-* ('regarding, in relation to': *pridiskutata* debated, *pridubata* remaining in doubt, *prilaborata* being worked on,

priskribata written about, *prizorgata* cared for). On the other hand, the passive perfective *-ita* has a similarly distinctive phraseology:

-ita-

1 *Al tio estas celitaj miaj planoj...*	It was to this that my plans were aimed.
2 *La atento de la tie starantoj estis direktita al alia veturilo.*	The attention of those standing by was directed to another vehicle.
3 *la partio firme marŝis laŭ la elektita vojo.*	The party firmly marched along the chosen path.
4 *Estis enkondukita nova ero.*	A new item was introduced.
5 *Multo estis farita en tiu antaŭŝtorma tempo.*	Much was done in the lull before the storm.

Here the *–ita* form has a wider range of uses, including verbs denoting completion (including verbal or reporting processes presumably typical in journalism: *dirita* said, *citita* cited, *indikita* indicated, *menciita* mentioned) as well as verbs denoting organization, decision making and causation (*finita* finished, *bazita* based, *elektita* chosen, *aranĝita* arranged, *atingita* achieved, *akceptita* accepted, *aprobita* approved, *planita* planned, *fiksita* fixed, *konstruita* built, *fondita* founded *kaŭzita* caused, *kreita* created). The form is also commonly used with mental-process verbs prefixed by *el-* (*ellaborita* elaborated, *elmetita* emitted, *elpensita* thought out, *elsendita* sent out, *eltrovita* discovered) as well as material-process verbs *for-* (*forlasita* forsaken, *forstrekita* stuck off, *forpuŝita* pushed away, *forpelita* driven off). The idea that Esperanto has classes of verbs has not surprisingly been opposed by some purists. Nevertheless, the assymmetrical correlation between grammatical form and semantic field has been suggested as a major organizational feature of language by Halliday (1994) in his discussion of 'marked' and 'unmarked' systems, a phenomenon we point out elsewhere in terms fo phraseology. In general, these patterns suggest *ata / ita* is not simply a symmetrical grammatical distinction to be applied to any lexical item, but a function of a choice in which one form or the other is either typical or marked. Thus the choice of a perfective for a verb of duration such as *konita* is a marked choice ('once well-known, known but no more'), as is the choice of a progressive for a typical verb of result: *konstruata* ('still under construction'). Rather than seeing this as an irregular feature of the language, Esperanto appears to have formed consistent phraseological patterns in more subtle ways than one would ordinarily suspect. However, this does contravene the usual way of seeing oppositions such as *ata / ita*, it simply imples that the presence of grammatical oppositions is not always a question of free choice but depends on lexical constraints.

Another important area of grammar that is particular to Esperanto involves ergative expressions where the grammatical subject is seen as the intended object of the verb. (Here we use the term 'ergative' to refer to the verb form rather than the traditional usage as a nominal case). This involves an obligatory and explicit classification of verbs into transitive / intransitive, and causes problems for some learners since the classification is imposed to all verbs and can be seen to be largely conventional. Esperanto allows transitive verbs to be marked ergatively by *-iĝ-*:

'they closed the factory'	*ili fermis la fabrikon* (transitive use)
'the factory closed'	*la fabriko fermiĝis* (intransitive ergative use)
'they began the lesson'	*ili komencas la lecionon* (transitive use)
'the lesson began'	*la leciono komenciĝis* (intransitive ergative use)

Esperanto also permits causative forms. Thus verbs which have an intransitive classification in the dictionary may be used accusatively with the causative morpheme *-ig-*:

'the water boils'	*la akvo bolas*
'to boil water'	*boligi akvon.*
'the baby sits in the chair'	*la bebo sidas en la seĝo*
'to sit the baby in the chair'	*sidigi la bebon en la seĝo.*

The initial choice of transitivity is fixed, and must effectively be learnt for each verb, for example *disvolvi* 'to develop' is transitive, whereas *evolui* is intransitive. While this distinction is consistent in Esperanto, this differs markedly with languages such as English and French which may use the same verbs transitively and intransitively.

3) Adverbs (*Fundamento* Rule 7).

The morpheme for lexical (or derived) adverbs is *-e*. No words are listed as initially open-class adverbs in the dictionaries, although some forms may occur more frequently as adverbs in practice (*certe* certainly, *fine* finally). As noted above, adverbs have a greater range of use in Esperanto, and it has been widely noted that their role as modifier is increasingly preferred over equivalent prepositional phrases (Kalocsay 1962, Piron 1989a). In the corpus *e*-forms as sentence adverbs are frequently used to signal textual conjunction (the following examples demonstrate

how adverbs have come to replace prepositional phrases: *efektive* in fact, *ekzemple* for example, *ĝenerale* in general). Forms in –*e* are also used as intensifiers of nouns or attributes (*multe* many, *entute* completely, *ĝuste* exactly, *maksimume* maximally). Certain compounds also tend to be used more commonly as adverbs and can be seen as idiomatic when translated literally (*verŝajne* = true + seemingly 'presumably', *unuavice* = first + in line 'in the first instance', *surbaze* on + basically 'on this basis', *delonge* from + longly 'for a long time', *ankoraŭfoje* still + at a time 'once again'). These forms are at times sentence adverbs with a textual function and are more frequent in the corpus than the prototypical adverbs demonstrated in traditional grammars (*rapide, malrapide, elokvente* quickly, slowly, eloquently etc.).

Lexical adverbs can take the accusative ending to indicate direction: *li laboras hejme*, he works at home, *li iras hejmen* he goes home (to his home), *li veturas norden* he is driving North(wards). In the corpus, 4% of *e*-forms are used with the accusative, and their use is typically restricted to expressions denoting spatial relations. There are also a small number of uses as modifiers of noun phrases (a number of differences between adverbs in Esperanto and English uses can be seen in the following: *kiam ni fanfarone iradis ĉirkaŭe kun altenlevita nazo*: 'when we proudly+ADV paraded around+ADV, nose pushed upwards+ADV+ACC'). Most -*en*-forms are applied to prepositions or lexical items denoting directions (*antaŭen*: forwards, *alten* up towards, *dekstren* to the right, *flanken* on the side, *suden* to the south). Other forms are restricted to a small set of lexical items: in the corpus we find *hejmen* 'homewards' (40 instances) and one instance each of *ĉielen* 'skywards', *indeksen* 'back to the index page', *renkonten* to the meeting. Potentially, the –*en* form is also used to paraphrase any prepositional phrase denoting direction (*hejmen* is thus used instead of *al la hejmo*), but this does not appear to be a widely exploited feature in the corpus.

Grammatical adverbs have no characteristic morpheme, and include many monosyllabic words: *nun* (now)*, ĵus* (just), *tuj* (immediately). Rather unusually, these are often modified by the -*e* form and other word class morphemes to form new adverbs and other classes of word (for example the corpus contains examples such as *nure*, perhaps translatable as an emphatic 'merely'). Other functional items such as *jes* 'yes' regularly take grammatical and other morphology (*jese* affirmatively, *jesi* to agree, to say yes). The series ending in –*aŭ* sometimes also takes adverbial and other

endings in this way (the corpus contains several instances of *adiaŭe* 'by way of farewell', *antstataŭe* 'as a replacement'). Further notes on the range and use of adverbs are made in the section on syntax, below.

2.4 Derivational Morphology.

The principles of word formation in Esperanto are extremely productive and were established by Rule 11: 'Compound words are formed by a junction of roots, the head word at the end'. This section attempts to establish to what extent the relatively free morphology of Esperanto is governed by underlying principles. Two basic systems were originally built into the language: firstly word formation by lexical morphemes (the suffixes and prefixes) and by functional morphemes (denoting nouns *-o*, adjectives *-a*, verbs *-i*). The core principle of agglutinating word formation (juxtaposition of morphemes and words) is that each word may be broken into minimal units of meaning which are stable and unchanging. For example, the word *senlaboreco* 'unemployment' consists of four morphemes:

SEN	*LABOR*	*EC*	*O*
'without' (prepositional **prefix**)	'work' (nominal **root**)	'quality' (abstract quality **suffix**)	'noun' (noun **morpheme**)

The consistency of derivational morphology means that there is virtually no limit to the application of a morpheme in Esperanto. For example, *-eco* is used to derive nominals from adjectives:

Esperanto		**English**		**French.**	
alta	-> *alteco*	high	-> h**eight**	haut	-> haut**eur**
amika	-> *amikeco*	friendly-> friend**ship**		amical	-> ami**tié**
certa	-> *certeco*	certain -> certain**ty**		certain	-> cert**itude**
granda	-> *grandeco*	large, big -> **size**		grand	-> grand**eur**
patra	-> *patreco*	fatherly -> father**hood**		paternel	-> patern**ité**

The potential application of *–ec-* to all adjectives in Esperanto clearly represents a considerable simplification of the forms derived from different historical processes in

English and French. The following examples demonstrate further the two basic processes of word formation in Esperanto: functional derivation (1: by changing the word class morpheme) and morphological derivation (2: by using a derivational suffix, the two main examples being *-ec* from adjectives and *-ad* from verbs):

Process	Dictionary Form	Derivations
Verb -> noun	*celi* 'to aim'	1 -> *celo* 'a target'
		2 -> *celado* 'targeting'
Verb-> adjective	*vivi* 'to live'	1 -> *viva* 'alive'
		2 -> *viveca* 'vivacious'
Adjective-> verb	*plena* 'full'	1 -> *pleni* 'to be full'
		2 -> *plenigi* 'to fill'
Adjective-> noun	*originala* 'original'	1 -> *originalo* 'an original'
		2 -> *originaleco* 'originality'
Noun -> adjective	*frato* 'brother'	1 -> *frata* 'brotherly'
		2 -> *frateca* 'fraternal'
Noun -> verb	*kandidato* 'candidate'	1 -> *kandidati* 'to stand for election'
		2 -> *kandidatiĝi* 'to be nominated as candidate'

Kalocsay and Waringhien (1985:393-522) argued that these patterns constitute a major advantage of Esperanto over rival systems such as Ido and Interlingua. This morphological system is highly specific to Esperanto and raises a number of theoretical issues relating to agglutination that are simply not evident in inflected or isolating languages, namely the range of derivational paradigms, the morpheme effect, and the issue of serial and parallel compounds.

1) Derivational paradigms.

Compounds involving lexical morphemes make up the majority of commonly occurring words in our text corpus, over and above word compounds. All lexical morphemes have a number of properties in common. It is impossible, for example, to insert functional morphemes between lexical items and lexical morphemes. Thus the lexical compound *ĉokoladskatolo* 'chocolate + box' can be written *ĉokolad<u>o</u>skatolo* with no change in meaning. But the regularly constructed compound *ĉokoladujo* 'chocolate box' (chocolate + container) can not be interrupted with a nominal *-o* as in **ĉokolad<u>o</u>ujo*. Furthermore, it appears that lexical morphemes have a more

fundamental effect on the meaning of the compound than lexical items which are superficially similar. So although in isolation the lexical morpheme *ujo* means 'container', the meaning of *ĉokoladujo* is very specific (*-uj-* as a lexical morpheme rather than a separate word implies 'unique container' rather than just 'container'). On the other hand *ĉokoladskatolo* can mean 'box for chocolate' or 'box made out of chocolate' where the exact interpretation is left to context and common sense (as with a number of serial compounds mentioned above). The distinction is just as marked when we compare other near-synonyms, for example: *loko* (a lexical item: 'place') and *–ej-* (a lexical morpheme for 'specific place'). The forms *trinkoloko, trinkloko* ('drinking place, watering hole') are possible, but only *trinkejo* - and not **trinkoejo* - can exist. Furthermore the word *trinkejo* (as with *ĉokoladujo, lernejo* 'school' and other derived words), specifically and conventionally means 'bar' (café is *kafejo*). Perhaps surprisingly, lexical morphemes can be seen to form very specific words, with more or less fixed conventional meanings. Lexical morphemes therefore deserve a different lexical status to functional morphemes on the one hand (word class items such as *-o* and *-int*) and lexical items or roots on the other.

Although there is no theoretical limit to series of morphemes combined into one word in Esperanto, the longest combinations in the corpus tend to involve four items including the word class morpheme. Few of these involve combinations of more than two lexical root morphemes ('lexical items'). From a sample of 1460 nouns, we found 66 four-morpheme compounds (4.5%) and 254 three-morpheme compounds (17%: all other nouns are two-morpheme nouns: root + word class marker *-o*). Most of the longer examples include one or more derivational morphemes *ad* (de-verbal noun)*, ar* (collective)*, ec* (de-adjectival noun)*, em* (tendency), *ig* (causative):

mok-turment-ad-o (mock / torment / deverbal / noun: 'mockery')
administr-ant-ar-o (administrate- / person / group / noun: 'administrative service')
interes-it-ec-o (interest / -ed / quality / noun: 'general interest')
sci-vol-em-o (know / wish / tendency / noun: 'curiosity')
pli-profund-ig-o (more / deep / cause / noun: 'deep penetration')

The average number of derived word forms for any open-class root is ten. Below we demonstrate some of the most productive open-class words in the corpus including inflectional and derived forms (frequency and approximate translation in parentheses):

A Nominal Paradigm

(*lando* 'country, nation': 39 forms, excluding proper nouns e.g. Svislando-Switzerland)

alilanda (4, of another country)

alilandaj (1, of another country)

alilandano (2, foreign national)

alilandanoj (2, foreign nationals)

alilande (2, in other countries)

diverslanda (6, multinational)

eksterlanda (4, foreign)

eksterlandaj (13, foreign)

n (5, foreign acc.)

eksterlandanoj (6, outsiders)

eksterlanden (2, out of the country)

enlanda (2, within the country)

fabellando (1, land of fable)

fremdlando (3, foreign country)

hejmlando (5, homeland)

landa (63, national)

landaj (19, national pl.)

landajn (2, national pl. acc.)

landan (6, national acc.)

landanoj (2, nationals)

landaro (1, group of nations)

landekonomia (1, economic)

landfamaj (1, nationally famous)

landlima (1, bordering)

landlimoj (1, borders)

landlimo (2, border)

landlimon (1, border acc.)

lando (266, country)

landoj (283, national)

landojn (20, country)

landon (20, country acc.)

landskalaj (1, nation-wide pl.)

laŭlande (1, by country)

multlanda (1, multinational)

plurlanda (2, of many nations)

samlandano (4, fellow national)

samlandaninoj (1, fellow country-women)

tutlanda (5, country-wide)

A Verbal Paradigm

(*aĉeti* 'to buy'. 17 derived forms)

aĉetas (2, buys)

aĉetata (2, bought)

aĉetataj (1, bought pl.)

aĉetebla (1, buyable)

aĉeteblas (2, is buyable)

aĉeti (18, to buy)

aĉetindaj (1, worth buying)

aĉetis (14, bought)

aĉetkostoj (1, buying costs)

aĉeto (1, a buy)

aĉetu (3, buy…!)

aĉetokontrakto (1, purchasing contract)

antaŭaĉetintoj (1, prior-purchasers)

produktaĵaĉetoj (1, product purchases)

An Adjectival Paradigm

(aktiva: 'active' adjective, 38 derived forms)

aktiva (30, active)	aktivadis (1, was active)	aktivado (2, being active)
aktivadon (1, being active acc.)	aktivaj (15, active pl.)	aktivaĵoj (1, activations)
aktivaĵojn (1, activities acc.)	aktivajn (1 active acc. pl.)	aktivan (17, active acc.)
aktivaj (2, active pl.)	aktive (52, actively)	aktiveco (21, activeness)
aktivecoj (7, activities)	aktivecojn (3, activities acc. pl.)	aktivi (1, to be active)
aktivigi (4, to activate)	aktivigo (4, activation)	aktivigon (1, activation acc.)
aktiviĝas (1, becomes active)	aktiviĝi (1, to become active)	aktiviĝis (1, became active)
aktiviĝo (2, becoming active)	aktivis (3, was active)	aktivismo (2, activism)
aktivula (1, like an activist)	aktivularo (5, group of activists)	aktivulo (2, activist)
aktivuloj (6, activists)	aktivulojn (3, activists acc.)	neaktiva (5, inactive)
paperaktivado (1, paper-activity)	pliaktiviĝas (1, gets more active)	pliaktivigo (1, motivation)
radioaktiva (2, radio-active)	reaktivigo (2, re-activization)	reaktivis (5, resumed activity)
sapeaktivan (1, undermining)	sociaktivularo (1,active group)	

2) The Morpheme effect.

Although the cycle 'noun -> verb -> adjective -> noun' etc. can be continued indefinitely, this is constrained by the what is known as the morpheme effect (Schubert 1989b). This takes the form of a rule: if a derivational morpheme is used on a functionally derived form, then a category morpheme must be used (the main ones being *-ad-* for verbs, *-ec-* for adjectives). The morpheme effect imposes a consistent principle according to which the basic function of a root word must be taken into account in derivations (whether noun, verb or adjective). Thus *martelo* a hammer becomes *marteli* 'to hammer' as a regular functional derivation, but a morphological derivation based on the verb *marteli* must then use the categorising morpheme *–ad-* which maintains a verbal sense: *martelado*: a hammering, *martelada* 'like a hammering' (otherwise the nominal ending *–o* simply brings us back to the basic noun 'a hammer'). Similarly, *homo* 'a human' can be changed functionally to *homa* 'human', but any derivations on the basis of this must use *–ec-* as in *homeco* humanity, *homeca* of humanity, humane.

A consequence of the morpheme effect is that the functional morpheme *-o* can not be applied in all cases, as can be seen in the following examples:

originala original	-> *originalo* an original	but	*nova* new -> ? *novo* ?'a newness'
celi to aim	-> *celo* a target	but	*segi* to saw -> ? *sego* ?'a sawing'

In these cases, Esperanto prefers to use the categorical morphemes (*ad, aĵ, ec, ul*: giving *segado* the action of sawing, *noveco* newness, novelty). It could be argued that the form ?*sego* is the instrument 'a saw', but Esperanto prefers instead to impose the instrument morpheme –*il*- for nominals based on verbal roots (*segilo* a saw, *ŝovelilo* a shovel, *kombilo* a comb: the implication is that these are still all verbal roots). For nominal roots the –*il*- morpheme is seen as redundant (*martelo* a hammer, *broso* a brush: the morpheme effect is not invoked here because these are nominal roots). It is because of these specific lexical constraints that the categorical morphemes are considered to be a specific class of derivational morpheme, and have a special function relating to the morpheme effect.

Such nominal uses of adjectives exist in English and other languages but may take on different meanings. In English and French, for example, nominal adjectives are taken to mean people: 'the beautiful', 'les beaux'. In Romance languages, adjectival forms preceded by a determiner are interpreted as nominal propositions which are the equivalent of clauses in English and Esperanto (French 'l'important' / Spanish 'lo importante' = *kio gravas...*'what's important'). Similarly, Esperanto conventionally uses nominal adjective forms for languages (*la angla* English, *la hispana* Spanish). To arrive at a nominal adjective for 'people' Esperanto prefers to use another categorising morpheme –*ul*- meaning 'person' (*la beluloj* the beautiful people), In the case of adjectival and verbal roots, nominal endings tend to signal a result rather than an adjectival or verbal quality, although this depends on the semantics of the adjective or verb, thus: -*aĵ*- 'thing' for nominal results of verbs or adjectives: *novaĵo* (news, from the adjective *nova* new) and *segaĵo* (sawdust) from the verb *segi* (to saw).

The cyclical nature of the morpheme effect was pointed out early on in the development of the language, notably by René de Saussure (1910) and pursued by Kalocsay (1963) and others. The principle relies on the distinction between lexical

morphemes and lexical items (essentially: lexical items maintain their lexical function – noun, adjective, verb - whereas lexical morphemes do not). In the Esperanto dictionary all lexical items are defined as primarily of one of three word classes, although there are some officially sanctioned dual entries (*pilot-* is either a verb or a noun: *piloti* 'to pilot' vs. *piloto* 'a pilot') as well as numerous cases where the root has changed function or has been redefined over time (*floro* 'a flower' was originally verbal, *barbaro* 'a barbarian' was listed by Zamenhof as adjectival).

3) Serial and Parallel compounds.

The potentially unlimited use of compound roots in Esperanto has provided an area of intense debate for grammarians however, and the morpheme effect has not been universally accepted (an alternative view favours classless roots, as in Chinese: Janton 1991). The issue centers on the grammatical status of roots (or 'lexical items' as we have called them above). The problem has been particularly difficult in the area of word compounds, since several interpretations are sometimes possible. How is one to predictably interpret, for example, that *voĉdoni* should mean 'to give voice' or 'to vote' and not 'to give vociferously' while *fingromontri* 'finger + show' should mean 'to point out' rather than 'to show finger'?

On the basis of Saussure's analysis of the derivational system, Kalocsay (1963) claimed that two different processes were in operation in the language. These processes ultimately depend on the role of the elements within the word compound, in particular whether the first element is a complement or a modifier of the second element (for a definition of these terms, see syntax, below). In serial compounds the main word is modified by either nouns or adverbs which, generally speaking, can be reformulated by a prepositional phrase: *manĝoŝranko -> ŝranko por manĝaĵoj* cupboard for food 'pantry', *fingromontri -> montri per fingro*: 'to show by finger'). Parallel compounds on the other hand, involve complements or attributes of the main noun: as in *voĉdoni -> doni voĉon* 'to give a voice' or vote, *onidiroj -> oni diras* 'one says', the standard word for rumours, *bluokulo -> blua okulo*: 'blue eye', blue-eyes, a nickname). Parallel compounds can be reformulated as an adjectival, verbal or prepositional phrase (parallel compounds are glossed as 'derivation' by Schubert 1989:259). Some details of these two compound types are set out below.

Serial compounds are the most frequent kind of lexical derivation. They are 'serial' in that they are combined serially from left to right, the right hand term (the main term according to Rule 11) determining the function of the left hand items. Kalocsay establishes the following three principles of serial derivation:

1) *Root + Noun = Noun + Noun.* Any root prefixed to a nominal root is a <u>nominal</u> modifier, related by some prepositional phrase (Noun + Noun: *paperkorbo = korbo por papero* paper basket, *vaporŝipo* steam ship or *ŝipo pelita de vaporo* a ship propelled by steam, Verb + Noun *manĝhoro = horo de maĝado* time for eating, dinner time, *skribmaŝino* (typewriter) = *maŝino por skribado* machine for writing. Adjectival prefixes are rarer and can be interpreted as nominal modifiers: A + Noun: *orhorloĝo = horloĝo el oro* clock made of gold. More often, however they are treated as parallel compounds, such as *bluokulo* below.

2) *Root + Adjective = Noun + Adjective*. Any root prefixed to an adjectival root (*-a*) is a <u>nominal</u> modifier (Noun + Adjective: *vaporvarma = varma pro vaporo* warm with steam, *muŝoplena = plena de muŝoj* full of flies, *sabloflava = flava kiel sablo* yellow as sand, sandy-yellow, Verb + Adjective: *servopreta = preta por servo* ready for service. Adjective + Adjective compounds are rare, usually separated by a hyphen, and interpreted as adverbial modifiers as though the first element is a nominal: *neĝe-blanka*: snow white.

3) *Root + Verb = Adverb + Verb*. Any nominal or verbal root added to a verbal root (*-i*) is understood to be an <u>adverbial</u> modifier: Noun + Verb: *fingromontri = montri fingre* or *montri per fingro* 'show by finger' (not *to show a finger), *matenmanĝi = manĝi matene* to eat in the morning, to breakfast (not *to eat the morning). Verb + Verb: *finfari = fari fine* to do finally (as in: *kion vi finfaris*? What did you do in the end?, not as in *to do the end).

Parallel compounds are seen as exceptions to the main pattern, often involving somewhat idiomatic expressions. Kalocsay and Waringhien (1985:412-422) discuss a number of exceptions to the serial pattern, including compound words such as *senforta* (strengthless 'without strength'), *grandanima* ('broad minded', magnanimous),

bluokulo (a blue-eyes), *voĉdoni* ('give voice', to vote). Following the general rules as stated above, these can not be analyzed serially: **sen+o forta* (nominal interpretation)*, *grando anima* (nominal interpretation)*, *bluo okulo* (nominal interpretation)*, *doni voĉe / *doni per voĉo* (adverbial or modifier interpretation). In all of these cases, Kalocsay and Waringhien assumed that a different principle of construction was in operation. Instead these expressions must be interpreted as complements or attributes in independent phrases which are then reanalyzed as adjectives, nouns and verbs i.e.:

(*sen forto*) + *a*   (*blua okulo*) + *o*

(*granda animo*) + *a*   (*doni voĉon*) + *i*

Because of the initial analysis as a phrase, these are therefore known as parallel compounds, and the category appears to account for the large number of expressions such as preposition + root (*porinfana* 'for children', *perforto* 'by strength' = violence), as well as idiomatic expressions such as *voĉdoni* and *grandanima* which had been previously condemned by purists. Other 'idiomatic' examples include: *ĉef-* (*ĉefministro*: Prime Minister, *ĉefverko*: chef-d'oeuvre), *grand-* (*grandduko* Grand Duke*, grandkomerco* big industry), *plen-* (*plenbuŝo* a mouthful, *plentago* a whole day). Although Adjective root + Adjective are classed as serial compounds by Kalocsay, we consider them as a particularly frequent type of parallel derivation. The adjectival root prefixed to a verbal root (*-i*) is assumed to be an attributive complement rather than an adverb: (*altstari = stari alta* = to stand tall not **to stand in a tall way*, *plenŝtopi = ŝtopi plena* to stop full 'to stuff', not **to stop fully*). In addition, the complement-based view of derivation justifies the analysis of the *-iĝ-* and *–ig-* morphemes as parallel derivations: *grandiĝi = iĝi granda* 'to get big', to grow, *grandigi = igi granda* 'to make big', to enlarge. This supports the idea that *iĝ* and *ig* are themselves verbal in nature (regardless of functional morpheme), and serves to reinforce the categorization of derivational suffixes into a series of classes as presented below (section 2.4).

While the distinction between serial and parallel compounds is still a controversial one in some quarters, we believe that by seeing compounds in terms of modifier and complement structures we have been able to identify two divergent but also

complementary systems in the language. This is not a preferred solution for purists, but one which does group together a number of phenomena that would otherwise have been seen as incoherent. Another implication of the morpheme effect is that functional morpheme endings (*o, i, a*) are redundant and not required for lexical items (i.e. *nov-* is intrinsically an adjective regardless of functional ending). It is true that poetry, with the permission of rule 16, does indeed proceed by eliminating the final vowel of several words, relying on position in the clause and on lexical clues for clarity. However, this is still a very marked way of writing in Esperanto, and like other languages Esperanto exploits the redundancy of functional morphemes in a number of ways, not least of which as a way of differentiating lexical items from grammatical elements and by signaling the specific function of the word in the sentence. Thus lexical adjectives may be expressed with final –*a,* as attributes, or with the adverbial –*e* if they modify the clause. Similarly, adverbial –*e* replaces –*o* and whole prepositions (as mentioned above *per mano -> mane* by hand). The grammatical morphology of Esperanto is therefore just as essential to the syntactic expression of phrases as it is to the formation of vocabulary.

2.5 Affixes

At the heart of Zamenhof's language scheme, the derivational (or 'lexical') morphemes are often cited as the most powerful and original aspect of the language. Their range of use and their general principles of application profoundly affect the lexical and syntactic systems of the language. We have already suggested above that they enjoy a special grammatical status in the language, and most linguists consider them to be free morphemes (the main restriction on functional morphemes, as we stated above, is they rarely appear in compounds between lexical roots). Zamenhof's original scheme has proven to be relatively stable, as our statistics suggest below. However, many of the morphemes have taken on a life of their own as lexical items, and some are almost exclusively used in fixed or formulaic expressions. The 'official' affixes were soon joined by unofficial lexical and semi-lexical neologisms as well as a large number of prepositional prefixes of the type normally found in German and English. Generally speaking, while many of the original prefixes are relatively underused (with the notable exception of *mal-*), unofficial prefixes (including prepositions and forms derived from the correlatives) have expanded in use. Suffixes have seen an opposite

trend: official suffixes are very widely used but few of the unofficial suffixes have been fully integrated into the language. In terms of the evolutionary development of the system, even the most widely used morphemes have become associated with specialized clusters of lexical items, a concept developed in the section on phraseology, below.

1) Official prefixes

A description of each official prefix follows (as an indication of overall distribution, total frequencies of occurrence in the corpus are placed in parentheses):

Bo- (2) 'in-law' as in *bofrato* 'brother-in-law'. Derived from the French idiom 'beau' (*beau frère*).

Dis- (120) 'separation, dispersal' a common Latin morpheme e.g. *disvastigo* 'distribution'. Most uses are as an adverb *dise* (separately) and as a complex preposition *disde* (from between).

Ek- (60) 'sudden'. Most commonly used for two verbs *ekvidi* (to glimpse) and *ekstari* (to stand up). *Ek* is used in the complex preposition *ekde* meaning 'since' (36 instances) and as an interjection *ek!* Kalocsay and Waringhien (1985) see *ek* (as well as *re-* and the suffix *-ad*) as aspect markers attached to verbs, a regular and obligatory feature of Slavic verbal morphology. There are a number of such forms in the corpus: *ektuŝis* 'lightly touched' and *ekbrilis* 'shone out' and over half of these originate in the Russian *Soveta Esperantisto* part of the corpus.

Eks- (10) 'ex-, former' as in *eks-soldato* 'former soldier'. *Eks* is more commonly used as an adjective root (20 instances): *eksa Jugoslavio* 'the former Yugoslavia'.

Ge- (50) 'both sexes' as in *Gepatroj* 'parents'. Exclusively used in the corpus for family relations, the most common form being *gefratoj* 'siblings') where other forms tend to imply both sexes without the need for *ge-*. (*samlandanoj* fellow-nationals, *Francoj* French people, *kritikantoj* critics, *esperantistoj* etc.)

Mal- (3116) 'opposite' as in *malvera* 'untrue, false'. The form *mal* is derived from negations formed in French: *malfeliĉa* = malheureux, unhappy, *mallerta* maladroit, clumsy, *malpura* malpropre, unclean, *malafabla* 'malgracieux, ungracious' (Piron 1991:27). *Mal-* is a notorious morpheme, and critics have suggested that it mirrors the mechanisms of thought control portrayed in George Orwell's *Newspeak*. Less

ominously, *mal-* can be seen as a highly productive morpheme permitting a large number of motivated lexical forms (as opposed to synthetic forms such as bad / worse, big / small etc.). Analogous forms also exist in Creoles and in the ethnic languages, such as the prefix –un- in English which can be applied freely to form neologisms. Although not intended as a negative morpheme, *mal-* usually expresses the negative of two polar opposites (*malbona*: bad, *maljusta*: unjust*, malforta* weak) or logical opposites that are negative in most cultures (*malantaŭ* behind, *malalta* short, *malrapida* slow). Some forms do however have neutral or even conventionally positive meanings (*malferma*: open, *malpeza* light, *malarmado* disarmament). Speakers are aware of the conventional nature of many of the more common uses of *mal-*, and 'un-negatives' are often exploited for humor. The corpus contains a number of examples from prose dialogues: *malachan malvesperon* 'un-terrible un-evening' for 'good morning'. Some *mal-* forms are replaced by writers who prefer to use simpler forms or neologisms (the hesitation between *mal-* forms and neologisms is discussed under phraseology, below).

Mis- (35) 'bad, mis-' *miskompreno* 'misunderstanding'. *misa* is used to mean 'wrong' (*io estas misa* something is wrong, something is up).

Pra- (8) 'primal' *praarbaroj* 'primeval forests'. More commonly occurs as an adjective *praaj tempoj* 'ancient times'.

Re- (80) 're-, repetition, reverse' as in *renovigo* renovation, *revoki* to call back. Although potentially an aspectual marker for verbs of repeated duration, *re* is most commonly used in more general constructions which correspond to existing forms in English, French and the other major donor languages (*reagi* 'react', *rememori* 'to remember', *reformo* 'reform'). *Re* is also commonly used in adverb form *ree* 'continually', *reen* 'back(wards)'.

2) Unofficial prefixes

Ĉef- (86) 'chief' as in *ĉefurbo* capital city, *ĉefministro* Prime minister. Originally designating a 'chief', the semantics of the word have widened, the most frequent uses being: *ĉefa* main, *ĉefe* principally, chiefly.

Ĉia- (6) 'every kind of' (adj.). A correlative (see below). Used as a prefix for the following forms: *ĉiaspeca* of every kind, *ĉiamaniere* in every manner, in any case, *ĉiarimede* by any means. The first two items appear to have emerged as more nuanced

and morphologically more analytic versions of the 'official' correlatives *ĉia* (every kind) and *ĉiel* (in every way).

Ĉiu- (58) 'every, each one' (determiner, pronoun). A correlative (see below). Used very frequently in a variety of compound forms, some of which appear to be nuanced versions of other correlatives: *ĉiuflanka* on every side, *ĉiufoje* every time, *ĉiujare* every year (and other similar time expressions), *ĉiukaze* in each case, *ĉiuloke* in every place (possibly equivalent to *ĉie* everywhere), *ĉiurimede* by all means, *ĉiuspeca* of all kinds (possibly equivalent to *ĉiarimede*, *ĉiaspeca*).

Fi- (19) 'nasty'. Officially an interjection, *fi* occurs mostly as a prefix. Used infrequently in a variety of compounds (*fikomplezemo* terrible complacency, *fiaĵoj* damned things, *fia* disgusting).

Fuŝ- (18) 'poor, bad'. A pejorative prefix, possibly equivalent to 'official' *fi*. Derived from the lexical item: *fuŝi* to mess up.

Mem- (32) reflexive modifier meaning 'by itself, alone' Used almost exclusively in the three forms *memstara* independent, *memrigarda* self-regarding, *memkompreneble-* self-evidently. *Mem* as specifier is distinguished from the pronoun *si, sia* (one's self) which refers back to a subject in the clause.

Multe- (10) 'many'. Used in three words: *multekosta* expensive, *multepeza* very heavy, *mutlenombra* multifarious.

Plej- (25) 'most'. Used in three words: *plejofte* 'most often', *plejparte* for the most part, *plejmulto* the majority.

Pli- (35) 'more'. Used in a variety of contexts, mostly nominal: *pliaĝuloj* older people, *plialtigo* a rise, *plibonigo* improvement, *plimulto* the most part.

Retro- (21) 'backwards'. Introduced for a specific meaning already denoted by *re-* and used commonly in the expressions *retroiri*: to go backwards, *retrorigardi* to look back.

Sam- (60) 'same'. Used in a number of contexts, most frequently in *samtempe* at the same time, *samtempulo* contemporary, *samideano* fellow-ideologist.

Sin- (82) 'self'. Used where the reflexive pronoun is the complement of a verb root: *sindefendo-* self defense, *sindonema-* 'self-giving', selfless, self-sacrificing *singarde* defensively, guardedly. Also frequently used in the somewhat more idiomatic forms: *sinteno* attitude, posture, 'self-posing', *sindevigo* commitment 'self-obligation'.

Tia- (30) 'such' used in two forms *tiamaniere* in such a manner (more emphatic form of *tiel*: thus) and *tiaspecaj* such kinds of…(an emphatic form of *tia* such a…).

Tiu- (120) 'this' used in four frequent expressions: *tiumomente* at that moment, *tiutempe* at that time (both equivalent to *tiam* then), *tiusence* in that sense, *tiuokaze* in that case.

3) Modifying and prepositional prefixes

The 'official' prefixes are often cited as a characteristic aspect of Esperanto's morphology, but they are by far outnumbered by the use of prepositions and adverbs as prefixes, especially with verbs and their derivations. With verbs prepositional prefixes can be considered to be complements of the verb . The infinitive *meti* (to put) demonstrates a variety of possible forms, each of which can be reformulated as a prepositional phrase which is a complement of the verb:

...<u>demeti</u> mian puloveron... (to take off my pull-over)
…la verduloj povis <u>enmeti</u> tri proponojn (the Greens were able to submit ('put in') 3 proposals)
…<u>intermeti</u> min inter ĝin kaj la tornistron (to place myself between it and the haversack)
…<u>kontraŭmeti</u> al iu refutan argumenton (put a contrary argument to someone)
…la demandon <u>transmeti</u> al la sekva etapo (put the question back to the next stage)

Prepositional prefixes with nouns also involve the principle of parallel compound formation, i.e. the preposition is considered to be part of an embedded stem: *subtera*: = (*sub tero*)+*a* 'under ground'+*a* or 'subterranean'. Venture (1977) points out that the preposition *el-* (from, out) has been the most productive prepositional prefix and is generally preferred over a number of alternatives: *elteni* (to hold out vs. *trateni* to hold through), *elkalkuli* (to work out = *finkalkuli* 'to final-calculate'), *ellegi* (to read out = *laŭtlegi* to read aloud). Some prefixes consist of an adverb indicating space or direction , itself sometimes derived from a preposition: *antaŭeniri* (to go forwards), *foriri* (to go away), *malsupreniri* (to go down, descend). Modifying expressions can also be expressed as independent words adjacent to the verb with little or no change in meaning (*iri malsupren* to go downwards, *iri for* to go away) and this use is the same for the other official prefixes that happen to be adverbs (*ek-, re-*).

4) **Official suffixes**

Kalocsay and Waringhien (1985: 437-8) divide Esperanto's derivational suffixes into four categories:

Categorical suffixes:	*ad, aĵ, ec, ul*
Nominal suffixes:	*an, ar, ej, er, estr, id, il, in, ing, ism, ist, uj*
Adjectival suffixes:	*ebl, em, end, ind*
Verbal suffixes:	*ig, iĝ*

These categories correspond to the morphological behaviour of each, although they may be expressed as any word class with the appropriate word-class morpheme (see discussion of the morpheme effect in section 2.4 above). Some morphemes are uncategorized (these include: *aĉ, eg, et, um*). These have no intrinsic lexical function, and their function will effectively be determined by the root they are attached to (this *ĉevaleg-* 'giant horse' is a nominal, while *grandeg-* 'enormous' is adjectival.

-aĉ- (23) 'pejorative', e.g. *odoraĉoj* 'foul smells' (uncategorized). Derived from Latinate forms such as *–asse, -azzo* etc. and designed to represent inherently bad quality (*ĉevalaĉo*: a nag is thus distinguished from other pejorative forms: *fiĉevalo* damn horse!).

-ad- (3225) (categorical) morpheme meaning 'duration', 'action'. The morpheme's most common use in our corpus is with de-verbal nouns where some duration is understood but also a resultant action, the most common forms being *agado* (action)*, gvidado* (guidance)*, intertraktado* (dealing)*, konstruado* (building activity)*, movado* (a movement). The morpheme is not an obligatory element: most verbs can simply take *–o* (*uzo*: a usage) where *–ad-* signals a nominal representation of the process of the verb rather than the verb's result (*informo*: a piece of information, *informado*: the act of informing, *manĝo*: a meal, *manĝado*: the act of eating). As an aspectual morpheme for verbs, *-ad* is mostly used with the past tense (192 occurrences) equivalent to a 'past imperfect' in several languages (*paroladis* was speaking, *akriĝadis* was becoming sour). Unlike the other derivational morphemes in Esperanto, *-ad-* is rarely

used as a lexical item on its own, although the corpus contains some examples of its use with non-finite verbs (*Jesuo…adis festi ekzakte ĉiun sepan tagon* Jesus continued to celebrate every seventh day, *ne plu eblas adi nian rilaton* it is no longer possible to continue our relationship). This role is more frequently fulfilled by the verb *daŭri* (intransitive) or *daŭrigi* (transitive, causative), although in the corpus these forms are used almost exclusively with nominal or other complements rather than to introduce verbs (*daŭrigis la vojon* continued the journey, *la subpremado daŭris* the 'continual' oppression went on). The adverb *daŭre* 'continually' is also preferred (46 occurrences) to the possible *ade* 'continually' (1 occurrence). The morpheme *ad* appears to have found a fundamental role in the language, although in Zamenhof's own writing its use appears to have been very limited (Wackrill (1907) only identifies *ad* with de-verbal noun forms including *kantado* singing, *skribado* writing, *uzado* usage).

-aĵ- (1131) 'concrete result' (categorical) morpheme meaning 'thing'. Derived from – *age-* and similar forms in the Romance languages. Used to form de-adjectival nouns (*novaĵo* 'news', *malfacilaĵo* difficulty) and de-verbal nouns where the emphasis is on material process or result (*manĝo* a meal, *manĝaĵo* food, *esto* existence, *estaĵo* a being, *konstruo* an abstract construction, *konstruaĵo* a building). When used (infrequently) with nominal roots, *-aĵ* also tends to indicate a result or some derived product (*teatro* theatre, *teatraĵo*: play, *verso* poetry, *versaĵo* piece of verse, *periodo* a period, *periodaĵo* a periodical). As a lexical item *aĵo* is very frequently used to mean 'thing'.

-an- (542) 'member' (nominal). Used most often as an independent lexical item. As a morpheme –*an-* indicates institutional membership (*policano* police officer, *civitano* citizen, *estrarano* board member) or nationality (*nederlandano* Netherlander, *usonano* U.S. citizen). Also used in an idiomatic use of the word *lingvo* for members of speech communities (*anglalingvano* Anglophone, *samlingvanoj* speakers of the same language).

-ar- (2595) 'collection, group' (nominal). Most often used in conjunction with morphemes denoting people (*ul, ist* etc.). A plural or other paraphrase is usually needed to translate the derivation in English (*junularo* young people, *loĝantaro* inhabitants, *homaro* humanity, *registaro* 'body of rulers' i.e. government). Also used with a limited but highly frequent set of nouns where idioms or specific collocations would be necessary in English: *montaro* a mountain range, *hararo* head of hair, *ŝtuparo* stair case, *armilaro* arsenal, armory (collection of weapons). Although a

potential morpheme for collective nouns (a herd, a pride, a flock), this application is unexploited in the corpus.

-ebl- (1544) '-able' (adjectival). Used in a wide variety of contexts and in many multiple compounds, usually with verbs (*kompreneble* understandably, *fleksebla* flexible). There are a number of uses as a passive verbal suffix indicating potential (*haveblas* is available 'havable', *aĉeteblas* is purchasable 'buyable') as well as a range of frequent uses as a lexical item in conjunction with other functional morphemes (*eble* possibly, *ebleco* possibility, *ebla* possible, *eblas* is possible, *ebligas* makes possible).

-ec- (3290) 'quality, property', represented by a variety of forms in other languages. (categorical). Used to nominalize a wide variety of adjective roots (*libereco* freedom, *juneco* youth, *certeco* certainty) and also to form abstract nouns on the basis of a number of nominal roots (*personeco* personality, *amikeco* friendship, *kamaradeco* comradeship). Less frequently used as an adjective (177 occurrences) such as *frateca* 'fraternal', *soleca* 'solitary', *respondeca* 'responsible'. When used as an adjective, the form provides doublets which contrast specific and generic qualities: compare *leda* 'leather' (adj.) / *ledeca* 'leathery', *latina* 'Latin' (adj.) / *latineca* 'latinesque'. As a verbal suffix *-ec-* is exclusively used in the word *respondeci* 'to be responsible'.

-eg- (504) 'extreme (size)' (uncategorized). Formally equivalent to superlative forms in Italian and other languages (*-issimo*). Used frequently with adjectives (*bonega* great) as well as nouns denoting size (*montego* giant mountain). Also used as an independent item, usually as an intensifier adverb (*ege* extremely, *treege* most extremely).

-ej- (837) 'place' (nominal). Used lexically and morphologically in a number of basic lexical items, all verbs (*preĝejo* church, *trinkejo* bar, café, *vendejo* shop, *kuirejo* kitchen, *lernejo* school).

-em- (553) 'tendency' (adjectival). Used frequently as a lexical item and as a morpheme in a wide variety of contexts to form nouns and less frequently adjectives, usually on the basis of verbs (*dankemo* gratefulness, *iniciatemo* initiative *scivolema* curious, literally 'tending to want to know').

-end- (19) 'obligation', a specific use derived from the participle system of Latin (adjectival). As adjective it is occurs almost exclusively in two forms: *atingenda* which must be reached, *farenda* which must be done. This form generally implies a passive reading, as with the analogous morpheme *-ind-*. Despite being initially

designated as a morpheme, *end-* is however more frequently used as an independent impersonal verb (20 occurrences) as in: *endas gajni tempon* 'it is necessary to save time'.

-er- (19) 'unit' (nominal). Occurs frequently as a lexical item but appears restricted to three compounds in the corpus: *ĉeneroj* links of a chain, *moneroj* coins, *programeroj* items on a program.

-estr- (183) 'leader' (nominal) as in *ŝtatestro* head of state.

-et- (271) 'diminutive' (uncategorized) as in *rideto* smile (small laugh), *strateto* small street, alley.

-id- (30) 'offspring' (nominal) for animals, botanical sub-genera or biblical races (*ursido* bear cub, *la kapridoj* the goat family, *Abrahamidoj* children of Abraham) but, as with other morphemes of this type, *-id-* is used most frequently as a lexical item on its own simply meaning 'child'.

-ig- (2939: verbal *–igi* 1947, nominal *–igo* 992) (verbal). Causative morpheme. Used to form causative verbs from a wide variety of adjectives (*certigi* to make certain 'ensure', *klarigi* to make clear 'explain', *publikigi* to make public 'publicize', *pretigi* to make ready) as well as to render intransitive verbs transitive (*atentigi* to warn, *haltigi* to cause to stop, *mortigi* to kill, *starigi* to make standing 'to set up'). Nominal forms in *-igo* represent similar meanings to the verbal forms and prove to have less direct translation equivalents in English: *ĉesigo* halt, *efektivigo* rendering effective, *plifirmigo* firming up, *forigo* causing to be away 'dismissal', *realigo* realization). As with 'make' in English, *igi* is used on its own with predicative attributes which take no accusative: *igi prezojn pli flekseblaj* make prices (be) more flexible. Unlike other verbs, occurrences of *–igi* are more often expressed as infinitives, and the majority of forms are preceded by modality (i.e. after auxiliaries like *povas* can, *volas* wants to or introduced by *por* 'in order to'). This may be linked with *igi'*s general meaning of 'obligation' (i.e. where modality is often presented indirectly or in non-finite form). With transitive verbs, the causative is typically used to projects its complement (an indirect noun) as the subject of the main verb (*farigi al li la taskon = igi, ke li faru la taskon* to have him do the task). When applied to intransitive roots, it is possible to interpret the verb either as an active or passive participle. This can usually be demonstrated (and disambiguated) by breaking the verb down into *igi* + passive participle or active participle (functioning as attributive complement): *lavigi* = (1) *igi lavanta* to make someone wash or (2) *igi lavita* to cause something to be washed i.e.

to have something washed. As mentioned under grammatical morphology, above (section 2.3), *igi* is used to change the usually fixed transitivity of verbs and does not necessarily relate to the semantics of the verb alone. For example *daŭri* 'to continue' is intransitive, so 'to cause to continue' must be expressed: *daŭrigi*. When an implicit complement is understood, however, the causative *igi* form is preferred: *ŝi daŭrigis* she continued (she went on to say…). However, there is hesitation over which form of the verb to use when the complement of the verb is left unsaid, and both forms appear to be used with equal frequency in the corpus: *tiu danĝero daŭras resti* 'that danger continues to stay' (a seemingly transitive use but where *daŭri* appears to function as a modal verb) vs. *sed oni daŭrigas labori por prepari* 'but one continues to work in order to prepare…' (where the *daŭrigi* has a clear complement, i.e. the verb *labori*).

-iĝ- (2697: verbal *iĝi* 2125, nominal *iĝo* 567). Ergative morpheme (i.e. the subject of an ergative verb is an assumed object of the verb's transitive form. The term *mediala* is also used). Used widely to render transitive verbs intransitive (*baziĝi* to be based on, *finiĝi* to finish, *komenciĝi* to begin, *interesiĝi* to be interested, *ŝanĝiĝi* to change) and to represent verbal processes derived from adjectives (*altiĝi* to rise, *plifirmiĝi* to get firm, *proksimiĝi* to get near, *tediĝi* to be bored, *trankviliĝi* to calm down). The most frequent nominal forms tend to be derived from verbs: *disvolviĝo* unraveling (interpreted as 'development'), *moviĝo* a move, *naskiĝo* birth, *renkontiĝo* meeting, *vekiĝo* awakening. *-Iĝ-* is further discussed below under phraseology.

-il- (462) 'tool' (nominal) used with traditional conceptions of tools (many derived from verbs: *ŝovelilo* shovel, *ŝlosilo* key, *fosilo* spade, *sonorilo* a bell) or with forms denoting the prototypical instrument of an activity (*aviadilo* aeroplane, *armilo* armament, weapon).

-in- (621) 'female' (nominal) used to mark feminine forms (*edzino* wife, *fraŭlino* Miss, *kanbino* girl, *sinjorino* Mrs, *virino* woman). Feminine forms for professions are often marked where they can not be grammatically signaled in English (*trejnistino* trainer, *sekretariino* secretary, *ĉevalistino* rider, *ĉefredaktorino* editor-in-chief, *kuracistino* doctor). While it might be thought that gender marking would be in decline, there has been a greater tendency to mark female roles morphologically, a tendency related to recent systematic moves to mark professions and titles in gender-marked languages (French, Spanish etc.). *-In-* has been criticized by commentators for being a sexist form, re-imposing Indo-European features on a potentially gender-free language. Words such as *Esperantistino, kuracistino, heroino* imply that the unmarked forms

(*Esperantisto, kuracisto, heroo*) are masculine and that terms for women are therefore relegated to the status of derived terms. There are very few examples of male derivations on the basis of neutral terms. The prefix *vir-* as in *virŝafo* 'ram', *virbovo* 'bull' is sometimes cited as a counter-example, but instances of this are rare and *virsimioj* 'male monkeys' (in a scientific text) is the only example we find in the corpus. Furthermore, the usual human morphemes used in genderless words (*ul, ist* and the participles) receive no further morphology suggesting that they are to be read as male or neutral (*junulo* 'young man / person', *junulino* 'young woman' but not *junino* 'young-female'). The mixed gender form *gejunuloj* (by analogy with *gesinjoroj* 'Mr. and Mrs.') is used widely in the corpus, further implying that the form *junuloj* is masculine. In contrast, the word *homo* 'human, person' is gender-neutral (and consequently is not used with *-in-*) and is at times used in compounds in a similar distribution to *persono* 'person' (*senhoma* person-less, 'unmanned', *herohomoj* heroic people, *kunhomo*: 'with person', companion, in parallel with the usual word for companion *kunulo* 'with person'). *-In-* is sometimes used as a full lexical item, and *ino* has been claimed as an autonomous term for 'woman', although it has an extremely limited use (12 occurrences, mainly in the collocation *inoj kaj idoj* 'women and children').

-ind- (285) 'worthy' (adjectival), a quasi–passive morpheme related to Latin-derived *-end-*. The most frequent applications are adverbs (*bedaŭrinde* unfortunately, 'worthy of regret', *konsiderinde* considerably, 'worthy of consideration') although adjectives also provide a variety of infrequent forms (*fidinda* trustworthy, *honorinda* honorable). As with *-end-* and *-ebl-*, despite its status as derivational morpheme, *-ind-* is more frequently used as an impersonal verb: *indas mencii ke...* it is worth mentioning that or an independent adjective *inda* 'worthy'.

-ing- (2) 'recipient, holder' (nominal) as in *akvingo* water bottle. *-Ing-* only occurs as an independent word in the corpus meaning 'scabbard' (*metu la glavon en vian ingon*: put the sword in your scabbard).

-ism- (923) '-ism' (nominal). Occurs both as noun for ideologies and behaviors (*socialismo* socialism, *alkoholismo* alcoholism) but also as an adjective, where English –ist(ic) would be used instead: *socialisma* socialist(ic), *humanisma* humanist(ic).

-ist- (1064) '-ist' / 'professional' or 'adherent' (nominal) as in *komunisto* communist, *komercisto* businessman. The participle forms (*anto, into*) can be used to characterize

people in general, but *–ist-* is reserved for full-time or professional activities. Thus *kritikisto* professional critic, can be opposed to *kritikanto* critic, opponent.

-obl-, -on- (fraction)*, **-op-*** (group) are all designed to be used with numerals (*duobla* double, *duono* half, *duopo* duo) (uncategorized). Of these morphemes, *obl* 'multiple' is the only one with a noticeable distribution in the corpus: it often used alone as an adverb: *sep oble dek* (seven times ten) and in compounds such as *multobligi* (to multiply).

-uj- (134) 'container' (nominal) is mostly used idiomatically to form the names of old-world countries from a word denoting their inhabitants (*Anglujo* England, *Judujo* Judea) and in an infrequent scattering of words in its original concrete sense of 'specialized container' (*fridujo* refrigerator, *abelujo* beehive, *dosierujoj* file holders, *instrumentujo* toolbox, *monujo* wallet). The ending *–io* has been proposed to replace *–ujo* for names of countries, partly on esthetic grounds but also in analogy with other forms (*Azio* – Asia). In the corpus the reformed terms appear to be more generally accepted although there is still hesitation: *Anglio / Anglujo (0 / 4), Britio - Britujo (14 / 2), Ĉinio - Ĉinujo (7 / 0), Egiptio – Egiptujo (0 / 4), Francio -Francujo (40 / 3), Germanio – Germanujo (4 / 6) Hindio – Hindujo (1 / 4) Patrio - Patrujo (25 / 10)* (England, Britain, China, Egypt, France, Germany, India, the Fatherland (derived from *patro* father). The names of countries have always provided Esperanto with some of the most difficult political and linguistic difficulties in word formation because Esperanto's forms were based on 19th Century models. For example, the Indian linguist Dashgupta (personal communication) campaigned against the form *Hindio* (which clashed with *hindio* the Hindi language) and this form was progressively replaced in UEA documents by *Hinda Unio* (Indian Union), then by the Sanskrit form *Bharato* and finally *Barato*. Finally, as with most lexical morphemes, *ujo* competes with lexical words, as can be seen in the use of *lando* as a root, leading to the alternative forms: *Skotlando*, *Skotujo* for 'Scotland' (in equal distribution in the corpus).

-ul- (207) 'person' (categorical). Characterizes a person by using adjectives (*fremdulo* foreigner, *junulo* young person, *spertulo* expert) and nouns (*krimulo* criminal, *samtempulo* contemporary, *sportulo* sportsman). One term is formed with a preposition (*kunulo* with person 'companion'). *Ulo* is a frequent informal word meaning 'guy, fellow'. Feminine forms are used in compounds, but *ulino* is not attested.

-um- (213). (uncategorized). A neutral morpheme, designed to fill lexical gaps where other derived forms would lead to misunderstandings (thus *kruco-* a cross, *krucumi*: to crucify, *akvo-* water: *akvumi-* to water plants). In the corpus, the form *plenumi* (to fulfill) accounts for most uses. Other examples include *cerbumi* to wonder, to wrack ones' brains, *trafoliumi* to flick through a book, *mastrumi* to master or manage, *gustumi* to taste). *Umi* exists as an independent verb, and its informal meaning 'to wander, potter around' can be associated with a number of informal, one-off uses (*kunumi* to hang around together, *ĉirkaŭumi* to wander around).

5) Unofficial suffixes

A number of 'unofficial' lexical morphemes have been proposed since the *Fundamento*. We have mentioned the *–io* replacement for *-ujo* above, and Kalocsay and Waringhien list a number of other possible forms. Very few have become widely used and none have gained a presence in the learning materials for the language. Some were proposed for technical purposes (such as ***-iz-*** 'cause to be' as in 'carbonize') or to cover asymmetries in the use of the normal affixes (as with *–if-* a verbal suffix for 'cause to be' + nominal: as in English sanct<u>if</u>y 'make a saint' where '*sanktigi'* arguably implies an adjective interpretation: 'make saintly'). In addition, the development of technical morphology meant that some unofficial forms resemble morphemes used elsewhere in the system, the best known example being ***-it-*** meaning 'salt' (in chemistry) and also 'inflammation' (in medicine) where both forms clash with *–it-*, a passive participle. Although several unofficial morphemes have been attested, the following are the only ones to occur in the corpus and the low figures suggest the relatively conservative nature of suffixation in the language:

-esk- (1) '-like' (adjectival) *Raymond-Schwartz-eska* 'Raymond-Schwartzesque'. The equivalent adjectival *eca* more generally includes this sense, as in *printempeca klimato* 'Spring-like climate', *la viro estis statueca* 'the man was statue-like'.
-iv- (2) 'potential' (verbal) as a form of causative marker as in *kreiva* 'creative', to be distinguished from *–ebl-* (as in *kreebla* 'create-able') which tends to be interpreted as a passive form. Despite the potential for this morpheme, only two instances are found in the corpus, both of them nominal *rezonivo* 'power of reasoning' and *sensivo* 'sensitivity'.

-oz- (12) 'full' (adjectival), a popular morpheme in poetic writing. The corpus contains a handful of examples (*dornozaj*: full of thorns, *haroza*: hairy, *pluvoza* full of dust, *tekstozaj*: text-rich). However, *-plena* is usually used in an equivalent sense, and is considerably more frequent (*sencoplena* full of meaning, *akvoplena* full of water, *vivoplena* full of life etc.).

The following lexical items are sometimes also used as lexical morphemes (space precludes discussion of other possible forms such as: *-aspekt-* (resembling) *-hom-* (person) etc.):

-*hav*-(9) 'with-' from *havi* 'to have' as in *monhava familio* 'a family with money', 'a monied family'. Equivalent to *kun* + root (*kun mono* with money).

-*pov*- (2) 'ability' from *povi* 'to do' as in *flugipova*: capable of flight, *daŭripova* durable. Generally equivalent to –*iv*-, above.

-*riĉ*- (2) 'rich', from the adjective *riĉa* as in *kolorriĉa* 'colorful'. Generally equivalent to the proposed morpheme –*oz*-.

-*simil*- (3) '-like', from the adjective *simila* as in *homsimila* 'humanlike'.

2.6 The Correlatives.

The correlatives are functional compound words, mostly pronouns and adverbs not derived by the normal processes of word formation, but determined from a cross-referenced table of abstract concepts. Together with the functional morphemes, the correlatives constitute one of the few non-derived features of Esperanto's vocabulary, although Zamenhof is said to have based the system more or less directly on Lithuanian, a regular system which has equivalents in most Indo-European languages (especially the prefix patterns *ki-, ti-, neni-*). Their overall distribution in the corpus is as follows:

	Ĉi- Every, all	*Ki-* What	*I-* Some	*Neni-* No-	*Ti-* That
-o 'thing'	*Ĉio* (329) Everything	*Kio* (673) What	*Io* (204) Something	*Nenio* (163) Nothing	*Tio* (1770) That
-u 'one'	*Ĉiu* (1232) Everyone	*Kiu* (2578) Who	*Iu* (456) Someone	*Neniu* (235) No one	*Tiu* (2317) That one
-a '-kind'	*Ĉia* (37) Every kind	*Kia* (110) What kind	*Ia* (109) Some kind	*Nenia* (40) No kind of	*Tia* (369) That kind of
-es (genitive)	*Ĉies* (7) Everyone's	*Kies* (141) Whose	*Ies* (7) Some one's	*Nenies* (4) No one's	*Ties* (135) That one's
-al 'reason'	*Ĉial* (0) For every reason	*Kial* (150) Why	*Ial* (4) For some reason	*Nenial* (3) For no reason	*Tial* (143) Because
-am 'time'	*Ĉiam* (293) Always	*Kiam* (643) When	*Iam* (142) At some time	*Neniam* (208) Never	*Tiam* (311) Then
-el 'way'	*Ĉiel* (2) Every way	*Kiel* (1183) How	*Iel* (65) Somehow	*Neniel* (23) In no way	*Tiel* (499) Thus, so
-om 'quantity'	*Ĉiom* (0) All quantities	*Kiom* (105) How much	*Iom* (281) Some	*Neniom* (6) None	*Tiom* (181) So much

The similarity with Lithuanian, which is almost completely regular, can be seen from the following sequences: *kiam* kada (when), *tiam* tada (then), *ĉiam* visada (always), *iam* kažkada (sometimes), *neniam* niekada (never) and *kiel* kaip (as, which way), *tiel* taip (so, this way), *ĉiel* visaip (every way), *iel* kažkaip (some way), *neniel* niekaip (no way). The main irregular form in Lithuanian is 'ten' instead of *tur ('there').

In Esperanto these words are known as *tabelvortoj* (table-words). The individual elements can either be seen semantically as 'atomistic' units of meaning or grammatically as functional and lexical morphemes. The *–u* forms have varying functions. The high frequency of *kiu* and *tiu* can be explained by their frequent use as relative pronouns with specific references (as in *se vi havas tekston, kiun vi volas proponi...*If you have a text which you would like to propose). As pronouns *neniu* and *kiu* refer to people (*neniu kuraĝis malagnoski*: no one dared deny), while *tiu* and *ĉiu* are more commonly used as determiners within the noun phrase (*tiu plej malamata besto* that most hated of creatures). *Iu* is used in both roles with equal distribution. The *–o* forms are more frequently used as pronouns. *Kio* is a non-specific relative pronoun (*mi ne sciis, kion fari* I didn't know what to do, *li ekkomprenis, pri kio temis* he realized what it was about). The *–a* forms are all used as determiners. Esperanto has therefore a complex set of determiners, specifying the reference of the noun phrase. As determiners, the *-u* and *-a* forms are not easily distinguished (for example *estas neniu tablo*: there's no table vs. *estas nenia tablo* there's no table). The distinction may be explained in terms of reference: *iu tablo*: some table (emphasizing quantity: one among many) versus *ia tablo*: some kind of table (emphasis on quality: some type of table). The corpus suggests that as determiner *ia* tends to be used with abstract nouns (the most frequent being: *ombro* shade, *dubo* doubt, *truko* trick) whereas *iu* is used with countable concrete nouns (*punkto* point, *loko* place, *voĉo* voice).

Despite their special status as a schematic word set, the correlatives have been organically integrated into the rest of the language, in particular Esperanto's system of agglutination. Zamenhof himself compared them with other morphemes in the *Fundamento*. The prefixes behave as lexical morphemes and are at times incorporated as such in the system. The suffixes behave as functional morphemes (defining word class), although their application to the rest of the system is restricted. The use of *a, e, o, u* as correlative suffixes coincides of course with the use of these vowels as functional morphemes. The forms *–o* 'nominal concrete' and *–a* 'adjectival quality' are easily seen as equivalent to their usual use in the language, while *-u* 'deictic pronoun / determiner' is unlikely to cause confusion with the verbal imperative form *-u*. The form *–e*, though adverbial, is more problematic and represents a more restricted

meaning than the general adverbial –*e*. The adverbial endings (*-al-, -am-, -el-, -om-*) are entirely artificial and reserved exclusively for use with the correlative prefixes. There some small etymological correspondences: *iam* 'some time' corresponds to the Latin form 'jam' / 'now', (also giving *jam* 'already'), while the genitive –*es* is related to Germanic genitives. If correlatives are used with other suffixes, a correlative prefix is always retained (thus *kialoj* 'reasons', but not **aloj*). The only exception to this involves attempts to innovate in literary writing, and the corpus contains examples of lexical items with the genitive (all taken from one piece by Julio Baghy): *futures espero* (the hope of the future), *la moderna beletres aprecanto* the modern follower of 'belles-lettres'. More typically, many of the more frequent correlatives (especially *ĉia, ĉiu, tia*) are used as derivational prefixes (these are the 'unofficial prefixes' noted in the section above).

The word *ali-* 'other' (originally an adjective: *alia*), may have become the sixth correlative prefix. The form is not frequent but is used with almost all the correlative endings: *alio* (5 something else), *alies* (3 someone else's), *aliuj* (4 others), as well as with derivations associated with the other correlative prefixes (like *ĉiaspeca, ĉiokaze*): *aliulo(j)* 28 (other people), *alispeca* (4) (another kind), *aliokaze* (2) (in other cases). The form *alie* (30 occurrences: meaning 'otherwise') is a regular use of the adverbial morpheme –*e* rather than the correlative. Even though *alie* is entirely regular, it is sometimes perceived an exception in relation to the correlative uses of *ali-*.

3 SYNTAX.

Although the basic morphology and vocabulary of Esperanto were set out in explicit terms by Zamenhof, many areas of syntax were essentially left open to interpretation. Zamenhof spent many years in correspondence about the development of the language, and some of the issues discussed at the time are still a matter of debate today. Grammars of the language eventually emerged in such detailed studies as Kalocsay and Waringhien's *Plena Analiza Gramatiko*. The question remains as to what constitutes the 'dominant word order' in Esperanto, and to what extent Esperanto's syntax is respected by the majority of speakers. Our corpus evidence suggests that Esperanto's syntax is used consistently, and that the system cannot as a whole be described as Latinate, Germanic or Slavic. How could such a consensus or a dominant style be arrived at, given the diversity of speakers? Firstly, a tolerant stylistic consensus was consciously manufactured within the community itself (not least by Zamenhof and his close followers). Secondly, although there was often radical experimentation, this was tempered by a process of rationalization: innovations within the language followed the fairly predictable patterns we have suggested above in our discussion of Creoles (i.e. innovations tend to reinforce the universal preference for relatively analytical and motivated structures). This has meant that some innovations have remained peripheral relative to the mainstream system (as with predicative verbs, for example), while other features of grammar that were theoretically possible have not been fully exploited (e.g. the use of the accusative to replace prepositions, as in Rule 14). As with aspects of morphology, it can be seen that some features have been systematized so consistently in the language that they appear to be unique to Esperanto and at times override the previous systems derived from ethnic languages (for example, the modifier use of adverbs, the neutral preposition *je*, both discussed below). The study of syntax in Esperanto therefore raises issues of style and permits us to ask whether the syntax of the language is independent of the ethnic languages which are spoken by the majority of its users.

A number of issues in the syntax and morphology of Esperanto revolve around the question of complements and modifiers. Since these terms are used differently by different school of linguists, we can should clarify that by <u>complement</u> we mean any

dependent element of a phrase or clause, including direct object complements, complements of prepositions and attributive complements (usually adjectives). Conversely, a <u>modifier</u> is a grammatically optional and mobile element, with various subdivisions including: adjuncts[2], attributes and prepositional phrases functioning as adverbs. In *The teacher of geography with long hair,* the complement is *of geography* while the modifier is *with long hair.* Any movement of the complement is restricted, so we cannot say: *The teacher with long hair [of geography].* The difference is not signaled explicitly in English or Esperanto, but does have an impact on several grammatical systems and accounts for several important constraints in both languages. For example, in English the complement can be reformulated as a noun compound: *The [geography teacher] with long hair*, but Esperanto does not allow this **La geografio instruisto kun longa hararo*. We have seen in our discussion of the 'morpheme effect' in section 2.4, that serial compounds in Esperanto typically contain modifiers rather than complements. A number of other distinctions in Esperanto also depend on whether a phrase is complement or modifier, as noted below.

3.1 Word Order.

Word order in Esperanto is often described as 'free'. When our corpus is searched for any clauses that contain SVO (Subject Verb Object), we find the following distribution:

SVO (705)   *Esploro ebligos transiron*. Exploration will enable a crossing.
OSV (258)   *Tion mi komprenas*. That I understand.
SOV (51)   *Vi nenion povis fari*. You could do nothing.
OVS (27)   *La leganton helpas resumo*. The reader is helped by a summary.
VOS (3)   *Ofte trafas nin malĉeesto*. Often absence strikes us.
VSO (1)   *Legas ni frazon plena de seniluziiĝo* We read a sentence full of disillusion.

This is rather limited search: we simply ask the computer to find, for example, only those sentences with an SVO or an SOV orde, and clearly in an authentic text corpus there are fewer of these than in a learner's textbook. Nevertheless, of the 1045 sentences with just three elements in the corpus, SVO is clearly the predominant

---

[2] In Hallidayan grammar an adjunct (Ab) is a modifier at the level of clause (usually circumstantial adverbs), while an attribute is a modifier at the level of the phrase (usually adjectives). All symbols

order, representing unmarked, indicative sentences. The SOV order is similar, tending to occur with sequences of pronouns before the verb (*mi ŝin alparoladis* I was talking to her). This is a marked choice, since there are significantly more occurrences of object pronouns in the 'normal' SVO position (*mi vidis lin*: I saw him). The VSO and VOS orders are notably rare. Their use is highly marked and most examples come from poetry or fiction (they are also associated with adjuncts (circumstantial prepositional phrases) as in: *El tiuj okuloj rigardis lin io* Out of those eyes, something was looking at him …). Word orders where the complement and other elements are placed initially can be accounted for by different sentence functions. OSV is an unmarked order in the language, usually corresponding to interrogatives (*Kion vi opinias?* What do you think?). It is also used when the complement is preferred as theme of the sentence (often where English would express a passive). In addition, OSV sentences are usually formed with pronoun subjects:

Ob	Sb	Ab	Vb
Ĉi tiun scenon	*oni*	*baldaŭ*	*prezentos*
[This scene-ACC	one	soon	will present]

'This scene will soon be presented.'

As mentioned above, all interrogatives involve the insertion of a question element (a function reserved for the *ki*-correlatives: *Kiom...* how much? *kie...* where? and certain derived forms *kiamaniere*: in what way? etc.). Although no other grammatical changes are required, subject-verb inversion usually takes place after instances of *kio* what?, *kiu* who? and *kies* whose?. In contrast, after adverbs such as *kiam* (when?) there is usually no change of word order. In polar interrogatives ('yes-no' questions), the particle *ĉu* must be inserted (position is usually initial, but when final *ĉu* functions as a question tag). Although no other grammatical change is required for polar interrogatives, in one third of the uses of *ĉu* OSV word order is preferred over SVO:

PCL	Ob	Sb	Vb
Ĉu	*gloramon*	*tio*	*montris*?
[(Interrogative particle)	glory-love-ACC	that	show]

'Did that show a lust for glory?'

used here can be found listed on the abbreviations page. Sentence functions are signaled with a small *b*

The OVS structure has a similar function to OSV: it makes the complement the theme of the clause. However, it is not used for interrogatives. Although relatively rare and therefore more marked than the OSV structure, examples of OVS are found in a variety of contexts:

Ob	Vb	Sb	Ab
Pli grandan minacon	*kreas*	*ĝi*	*por homa civilizacio*
[A greater menace-ACC	creates	it	for human civilization]

'It creates a greater menace for human civilization'.

Ab	Ob	Vb	Ab	Sb
Antaŭ kelka tempo	*sin*	*trovis*	*en ia gazeto*	*sekvanta anonco*
[Some time ago	itself-ACC	found	in some magazine	following notice]

'Some time ago the following notice appeared in a magazine:'

Ob	Ab	Vb	Sb
Lin	*laste por nokto*	*salutis*	*kampulo*
[him-ACC	lastly for night	greeted	a peasant]

'He was wished a good night by a peasant' [from a poem translated by Grabowski]

As shown in several examples above, adjuncts are more freely placed in Esperanto than they are in English, and are often found interrupting the normal SV or VO sequences. The word orders OVS and OSV rarely occur without adjuncts in prominent positions, and subject-verb inversion in other cases (i.e. simply AVS), is almost always accompanied by the introduction of circumstantial adjuncts. This process is common to English, French and other languages. In a sample analysis of 115 sentences from the corpus, AVS rather than SVO appears to be the most frequently occurring order in simple clauses in the corpus (AVS accounts for 10% of all clause complexes in the sample). Indeed, the analysis reveals that only a further 10% of the sentences in the corpus involve the typical sequence SVO, a useful corrective to the simplification that SVO is the 'dominant' word order. In terms of function, sentence initial adjuncts have a significant role to play in texts where there is a clear sequence of events or locations:

(Sb, Vb, Ob) to distinguish them from parts of speech.

Ab	Vb	Sb
Sube	*estas publikigitaj*	*la respondoj de Gorbaĉov.*
'Below	are reproduced (published)	Gorbachov's replies'

Ab	Vb	Sb
Kune kun la verdikto	*aperos*	*la nomoj de la juĝintoj*
'Together with the verdict	will appear	the names of the judges'

Ab	Vb	Sb
En lia voĉo	*aŭdeblas*	*surpriziĝo*
'In his voice	can be heard	surprise'

Ab	Vb	Sb	Ab
Tuj post la konferenco	*estis*	*akcepto*	*en la sama salono*
'Immediately after the conference	was	a reception	in the same room'

AVS generally coincides with verbs indicating appearance and abstract location (textual, temporal as well as spatial adjuncts: comparable with inversions in other European languages such as 'here lies…' in English, 'ici se trouve' in French). This has been posited as a factor affecting grammatical systems of voice, reflexivity and transitivity in other languages (adjuncts and VS inversion often coincide with specific choice of verb and verb form such as intransitives denoting location *troviĝas, montriĝas, vidiĝas* 'are found, are shown, can be seen', just in the same way that passives are often accompanied by 'by' + complement). In contrast, sentence adverbs such as *do* thus, *certe* certainly, *fakte* in fact etc. do not involve this kind of SV inversion (*Do, ĉio finiĝas bone*: So, everything ends well).

3.2 Clause Structure.

Clause structure in Esperanto is oriented around the verb: a finite verb is usually equivalent to a clause. The clause may function as modifier of another clause, in which case the position of the main clause and dependent clause are interchangeable[3]:

Main Clause

Vb	//		Cj	Sb	Vb	Ob
Helpas	*//*		*se*	*oni*	*konas*	*Esperanton*
(It) helps	//		if	one	knows	Esperanto.

                                          Main Clause

Cj	Ci	Vb	[Vb]	//	Vb	[Vb	Ob]
Se	*al vi*	*plaĉas*	*prelegi,*	*//*	*bonvolu*	*indiki*	*tion*
If	(to you)	(it) pleases	give a talk	//	please	indicate	that

('If you'd like to give a talk, please indicate so').

The clause can also function as complement of the verb:

Sb	Vb	[Cj	V	Ca]	
mi	*opinias*	*[ke*	*estas*	*tute*	*maljuste]*
I	think	[that	(it) is	completely	unfair]

The clause can also function as complement of the noun. In many cases this involves a noun complement clause that takes the same form as the verb complement clause above. This formulation often does not have a direct equivalent in English:

Sb	Vb	Ob	[Cj	Sb	Vb	Ob]
li	*montris*	*sian dankemon*	*[ke*	*ili*	*helpis*	*lin]*
he	showed	his gratitude	[that	they	had helped	him]

'he showed his gratitude for their help'

Clauses can also be complements of other elements, for example adjectives:

---

[3] Double obliques // are a Hallidayan (1994) convention indicating a modifying clause. Square brackets [ ] indicate that the clause is a rankshifted or relative clause (i.e. the clause functions as subject, complement or attribute).

Vb	Ca	[Cj	Sb	Vb	Ca]
ne estas	*sufiĉe*	[*ke*	*nia vivo*	*estas*	*en danĝero*]
it's not	enough	[that	our life is	in danger]	

Clauses may be complements of prepositional phrases, a frequent feature of Esperanto (English has no equivalent of expressions such as *pri tio ke, pro tio ke*):

Sb	Vb	Ci	[Cj	S	Vb	Ca]]
mi	*konscias*	*pri tio*	[*ke*	*tiu okupo*	*estas*	*supraĵa*]
I	am aware	about (the fact)	[that	this occupation	is	superficial].

In relative clauses, the clause functions as an attribute of the noun and is linked by a pronoun rather than a conjunction. A pronoun in the higher clause may or may not be deleted (it is often deleted in English and Esperanto):

Sb	Vb	Ob	[Ob	Sb	Vb]
Mi	*scias*	(-)	[*kion*	*ili*	*faras*]
I	know	(-)	[what	they	are doing]

It is possible however to state the higher clause noun in Esperanto, although this is a marked form in English:

Sb	Vb	Ob	[Ob	Sb	Vb]
Li	*vidis*	*tion,*	*kion*	*li*	*serĉis*
I	saw	(that)	[(which) he		was looking for]

'He saw what he was looking for'

Participles may also form non-finite relative clauses as attributes, a function that corresponds to their adjectival ending (*-anta, -inta, -ata, -ita* etc.). Esperanto permits these to be placed before or after the noun, and also between the subject and its active verb. Attributes therefore behave just as modifiers do in terms of movement and appear to have more freedom of placement than in English:

Cj	Sb	[Ca	Ob]	Vb
Kaj	*krio*	[*skuanta*	*aeron*]	*trakuris.*
And	a cry	[shaking	the air ACC]	ran though.

'And a cry ran through, shaking the air'

Since relative-clause participles are considered attributes, they must agree for case with their higher referent, as in the following example:

Sb	Vb	Ob	[A	Ca	Ce]
La festenestro	*gustumis*	*la akvon*	[*nun*	*fariĝintan*	*vino*]
The party-master	tasted	the water	[now	become-ACC	wine].

These can be seen to be functionally equivalent to 'full' relative clauses: *la krio* [(*skuanta / kiu skuas*) *la aeron,*] 'the cry [(shaking / which shakes) the air]'. But participles are not limited to a role as attribute, they can also be complements and modifiers (sentence adjuncts). In the examples below, the participle phrase in (1) is an attribute of the noun while in (2) it is a complement of the verb. This morphological distinction in Esperanto is expressed in English by different word orders or by paraphrase:

(1) *Ili vidis la kamaradojn[falantajn]* = They saw the [falling] comrades.
  (*Ili vidis la kamaradojn, kiuj falas* = They saw the comrades [who were falling])

(2) *Ili vidis la kamaradojn /falantaj* = They saw the comrades / falling.
 (*Ili vidis la kamaradojn // dum ili falas* = They saw the comrades$_i$ / as they$_i$ fell

This can be contrasted with a third option in Esperanto, which is to use the participle as a modifier of the clause, in which case it takes an adverbial morpheme (-*e*) and can be placed in free position in the clause as a whole:

(3) *Ili vidis la kamaradojn // falante* = They saw the comrades // falling.
   (*Falante, // ili vidis la kamaradojn*) = As they$_i$ fell, // they$_i$ saw the comrades$_j$.

Syntactically, the difference can be represented as levels of dependence (we simplify the X' projections here):

(1) *Ili vidis la kamaradojn falantajn.*

```
IP
|      \
NP     VP
|      |      \
Pro    V      NP
Ili    vidis  |      \
              D      N'
              la     |              \
                     N              AP
                     kamaradojn     A
                                    falantajn
```

(2) *Ili vidis la kamaradojn falantaj.*

```
IP
|      \
NP     VP
|      |                              \
pro    V'                             AP
Ili    |      \                       A
       V      NP                      falantaj
       vidis  |      \
              D      N
              la     kamaradojn
```

(3) *Ili vidis la kamaradojn falante.*

```
IP
|                                     \
IP                                    |
|      \                              AdvP
NP     VP                             |
|      |      \                       Adv
Ili    V      NP                      |
       |      |      \                falante
       vidis  D      N
              la     kamaradojn
```
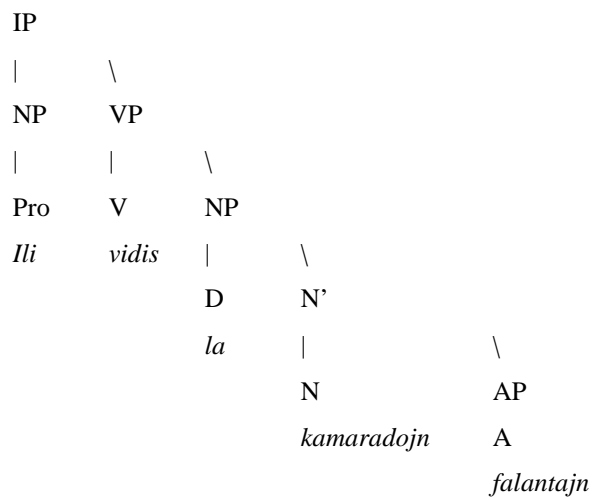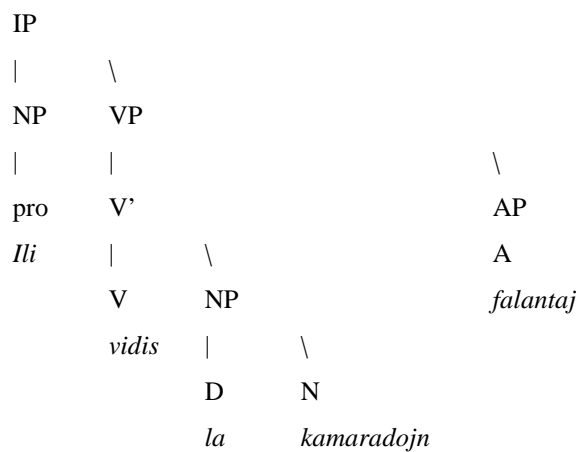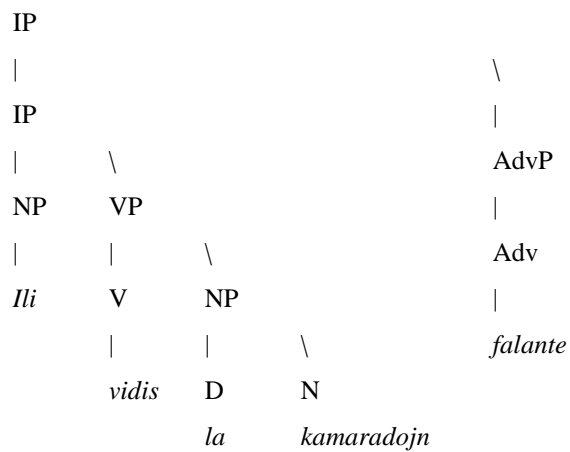
In other words, in (1) the adjective is a attributive modifier of the NP and bound to it (note the alternative wording within the NP: *ili trovis la <u>falantajn</u> kamaradojn*), whereas in (2) the AP is outside the NP structure, and it can be considered an attributive complement of the verb. The AP can thus be moved around within the VP: *ili vidis <u>falantaj</u> la kamaradojn*. It can be seen that only certain verbs allow this structure: *vidi, trovi, kredi, konsideri* (as in English: to see, to find, to believe, to consider), and Esperanto also commonly uses similar attributive complements after *igi* (to make, to cause) and *teni* (to hold). The attributive complement can be an adjective or a noun, in the nominative case: *ili supozis tio**n** la fino* (they supposed that (to be) the end), *la mortula mano tenis ĝi**n** rigida* (the dead-man's hand held it tight).

In (2), the reason that *falantaj* is still seen to unambiguously refer to the NP *kamaradojn* and not the subject (*ili*) is that the adverb participle *falant**e*** is conventionally used in Esperanto to refer to the subject. In sentence (3) then, the adverbial participle is independent of the VP, and positioned as an adjunct to the whole structure of IP. In other words, participles can also function as modifiers of the clause, in which case they must take an adverbial *–e*. The following example demonstrates a similar formulation, this time introducing a subordinate clause:

Ab	Sb	Ab	Vb	//	Vb	Ob
Diablon	*li*	*subite*	*ekpensis*	//	*<u>rememorante</u>*	*iam legitan libron.*
Damn	he	suddenly	thought,	//	remembering	a once-read book.

The adverbial morpheme is also obligatory for adjectives introduced by impersonal verb-complement clauses (*estas grave ke* it is important that*, estas necese ke* it is necessary that etc.). This convention is adhered to in the corpus, although it has not always been as widely used. Zamenhof (1962) at times used adverb forms with 'empty' pronouns (**ĝi estas neoportune*), although this has as much to do with the role of the pronoun as the role of the adverb morpheme. This use of adverbs may also be linked to the fact that active verbs are not always necessary in sentences. For example, the combination of particle *jen* 'here is' + participle functions in this way in the following example: *En la unua fasko da anekdotoj jen ĉiuj garantiite malnovaj*. 'In the first batch of stories, here (they are) all guaranteed to be old ones'. Subjects may also be considered optional, since the widespread use of impersonal verbs means that

subject-less verbs are commonly seen in sentence-initial position (*ŝajnas ke* it seems that, *aspektas ke* it looks as though, *montriĝis ke* it was shown that). Some complex sentences may lack formal subjects altogether:

Ab	Vb	Vb	Ab
Dum la nuklea epoko	*ne eblas*	[*vivi*	*almenaŭ ne longe*
In the nuclear age	(it) is not possible	[to live,	at least not for long,

Ci	(Ci -> cont.)
kun la pensmanierioj,	*kaj kutimoj de la epoko ŝtona.*]
with the thought-patterns	and habits of the stone age.]

Although identical to the European languages in many areas of syntax, we have seen that there are considerable differences of range and application of the general principles. For example, in English the clause can stand as subject of the verb: '[That he was frightened] does not surprise me'. But this is not a valid construction in Esperanto: *[Ke li timis] ne surprizas min*. In addition, many differences stem from different lexical restrictions on syntax (a point made more fully under phraseology, below). For example, in the case of noun complement clauses, not every language shares the same set of nouns which permit this construction. English permits a much wider set of nouns than French (compare *the idea that* -> l'idée que, the belief that -> la *conviction que*, but not *the thought that* -> *la pensée que, *the requirement that* -> *l'exigence que). Esperanto permits a similar set of nouns to English (*la espero ke* the hope that, *la ideo ke* the idea that, *la fakto ke* the fact that, *la timo ke* the fear that) but we saw above for *pro tio ke,* Esperanto also allows constructions that are structurally dissimilar to English. For example, *en la kazo ke* must be rendered by <u>to</u>-complementizers in English. Similarly, Esperanto seems to have a wider range of permissible '*ke* + active verb' clauses, where English requires 'to-' or '–ing' constructions: *estas do ŝia devo, ke ŝi oferu sian vivon...* 'it is her duty <u>to</u> sacrifice her life'. Nouns are often used in Esperanto where English simply attaches the that-clause to an adjective, as with particularly common expression *afero ke*: *Estas memkomprenebla afero* [*ke ili donos ekonomian evoluon*]: 'It's obvious [<u>that</u> they will create economic progress]'. *Afero* also frequently occurs in 'thing clauses', where the complement clause (always introduced by *esti*) is first classified as some kind of fact,

desire, observation: *interesa afero estas* [*ke en Svedio oni multe parolas pri mediprotektado*] an interesting thing is [that one often talks in Sweden about environmental protection]'). More generally, the conjunction in these clauses in obligatory in Esperanto, whereas it is not in English:

Sb	Vb	Ob	[Cj	Sb	Vb	Ob	Ab]
Kamila havis	*la senton*		[*ke*	*ŝi*	*rigardis*	*lin*	*el malproksimo*]
Kamila had	the feeling		[(that) she		was looking	at him	from afar].

This is similarly true of verb complement clauses:

Sb	Vb	Cj	Sb	Vb	Ci
Li	*diris*	[*ke*	*la poemo*	*celas*	*al li*]
He	said	[(that) the poem		is aimed	at him.]

Another divergence involves the interaction between syntax and word form, in particular the effect of syntactic structure on tenses (verb concord) and other grammatical features of the verb. We have noted the use of the volative *-u* in independent clauses, but it is also used obligatorily (or at least, consistently) in all dependent clauses introduced by *tiel ke* (such that)*, por ke* (in order that, so that) and in the expression *pri tio ke* 'about that': *la partio zorgos pri tio ke estu elektitaj la plej indaj homoj* 'The party will see to it that the most suitable people (will) be elected'. The future tense, in Esperanto as in French and other languages is usually marked explicitly in the subordinate clause where it is not in English: *Kiam ni havos ion por prezenti al ili* (Future) *// estos interese* (Future) 'When we have something to present to them (Present) // it will be interesting' (Future). The conditional tense is similarly used in hypothetical clauses introduced by *se* 'if' where in English a past tense (originally a subjunctive) or in French the imperfect past tense would be used: *se mi estus kelkajn jarojn pli juna* 'if I were several years younger', *se vi povus vi kondamnus la tutan mondon* 'If you could (were able to) you would condemn the whole world'.

Tense concord is not an obligatory feature elsewhere in Esperanto however, and dependent complement clauses often have divergent tenses. For example, in English

and French a past tense main clause shifts the tense in the dependent clause to a form of the past tense, whereas in Esperanto the tense chosen is understood in terms of its aspectual relation to the main clause:

1a) *Oni decidis*   *ke oni faros feston* (Future: Prospective)

1b) It was decided   that there would be a party (Conditional or Future-in-the-Past)

2a) *Vi legis*   *ke membro P.H. kreas hejmpaĝon* (Present: Progressive)

2b) You (have) read   that member P.H. was creating a home page (Past Progressive)

3a) *Oni malkovris*   *ke ŝi estis altnivela kontisto* (Past: Perfective)

3b) They discovered   that she had been a high-level accountant (Past: Perfective)

The aspectual relation 'progressive' is also maintained in conjunction with expressions denoting a certain length of time, so that in modifying clauses a present tense is used where English requires a perfective form, as in *ili manĝis* (Simple Past) *nenion depost la frua mateno* 'they has not eaten (Past Perfective) since the early morning'.

3.3 Noun Phrases.

The general structure of the NP is similar to that of English (see grammatical morphology for discussion of the syntax of accusative *–n*), except that complements and modifiers can theoretically precede or follow the head noun (the unmarked position is for APs to precede and for PPs to follow):

NP-> (D) (Num.) (AP complement or modifier) N (AP complement or modifier).

1) Determiners in the noun phrase:

La (invariable article) stands in contrast with either zero (*homo* a person) or other determiners (*tiu viro* that man, *kiu viro*: what man?). The relative difficulty of *la* has often been noted, and Zamenhof recognized that the learners who speak Slavic languages might have problems with it. The traditional explanation is that *la* is used to refer to a generic concept (in English this is not always the case. Compare: *la estonteco* 'the future' with *la historio* '(zero) history') or concrete items of previous

discourse. However, in texts from the corpus *la* usually signals that a noun is modified either by an adjective or in the majority of cases by a prepositional phrase. The following are typical uses (all are generic uses rather than anaphoric):

...kiel <u>la</u> rajto <u>de</u> ĉiu civitano…   (…as the right of every citizen)
...el <u>la</u> vidpuntko <u>de</u> edukado...   (from the viewpoint of education)
...dum <u>la</u> transdono <u>de</u> la teksto...   (during the exchange of the text)
...transdonus <u>la</u> esencon <u>de</u> la afero...   (would get across the idea of the issue)
...eĉ <u>la</u> ideo <u>pri</u> tio...   (even the idea of it…)

However, 'zero-article' can also refer to post-modified generic concepts: *malĝusta estis mia ideo ke [Ø] <u>protekto de</u> pensa libereco eblas*: 'it was wrong, my idea that (the) defense of free thought is possible'. Generally speaking, *la* is used when the head noun is modified by a prepositional phrase (PP): *la + N + PP*, while zero-article is likely to be used where the PP is a complement not a modifier. Esperanto corresponds in this way to English, as in the prototypical example: 'a teacher of geography' (*instruisto de geografio*: [Ø] + N + PP + [Ø] + NP complement). Similarly, we can contrast the grammatical 'the teacher with the long hair' / *la instruisto kun la longa hararo* with the ungrammatical '*a teacher with the long hair' (**instruisto kun la longa hararo*: *[Ø] + N + PP + *la* + NP modifier), where the specific referent signaled by 'the' or *la* is negated by the principal determiner, which is zero / or 'a' in the case of the second ungrammatical examples. Apart from this restriction, the zero article is most consistently used for plurals: *homoj penas* 'people try', *problemoj abundas* 'problems abound' and after certain prepositions functioning as adjuncts (i.e. not within NPs). *Kun, per, pro* introduce zero-article more frequently than others: *kun netradukebla mieno* with inscrutable features, *per voĉo* by voice, *pro sekureco* because of security.

The opposition zero-article / article in Esperanto is roughly symmetrical: in the corpus, 650 out of 1400 sampled nouns are determined by *la*, with around 200 instances of the other more common determiners *tiu, nia, ilia* (this, our, their). Although there is no 'indefinite article' in the language, a range of devices are available to indicate vagueness: *kelkaj* several, *certa* a certain, *iu* some (as in *iuj politikistoj provis:* some politicians did try). These forms are considered determiners because they exclusively specify the noun: no other determiners can be used alongside

them. The correlatives *kia* 'what kind of', *ia* 'some kind of' (as in *spureto de ia latineca akĉento* trace of some kind of Latin accent) are also used in this way. The most common determiners (*tiu, kiu*) act as specifiers, signaling selection of reference between the noun phrase and others in the immediate context. Certain specifying particles can be also used before or after the determiner: *ĉi* (a proximate particle used with both pronouns and determiners: *ĉiu ĉi* everyone here, *ĉi tiu jaro* this year but also on its own: *ĉi-jara* this year's), *ajn* (a non-selective particle: *iu ajn lingvo* any language whatsoever), *mem* (signifying an emphatic independent 'self' *mi mem,* myself *la viro mem* the man himself) and *eĉ* ('even' an emphatic particle signaling exception *eĉ mi* even I). All of these particles typically require the presence of a determiner (they are not specifiers in themselves). *Ajn* is said to be limited to uses with *kiu* and *iu* (as defined by Zamenhof in the *Ekzercaro* and by Venture: 1977) but the corpus contains uses with all the correlative prefixes (*ĉiaj ajn* any whatsoever, *neniu ajn* no one whatsoever).

In addition, words combining with the preposition *da* form quantifiers (a specific type of determiner) in expressions of quantity: *multe da ombro* much shade, *pli da informoj* more information, *milionoj da Sovetanoj* millions of Soviet citizens. This is especially associated with the correlative words ending in *–om* (*kiom da tempo?* How much time? *iom da dubo* some doubt). Since *da*-expressions function as qualifying elements in the phrase, *da* can not used in conjunction with other determiners (*iom <u>de la</u> kultura ĉirkaŭaĵo* 'some of the cultural surroundings', not **iom da la…*) and with pronouns, a characteristic common to all determiners: *multe <u>de tiuj</u>, kiuj laboris tie* many <u>of those</u> who worked there. However, because *da* is conventionally considered to be a preposition, the right hand noun (which is semantically the principal element) does not take the accusative case even where the quantifying expression is an adverb (*ŝi atentigis multe da homoj*: she warned lots of people). Nevertheless, the intuition that these form part of the noun phrase rather than a prepositional phrase is represented in a handful of errors in the corpus (*se ili iam ricevos sufiĉe da *influon*: if they ever get any influence-ACC.). The distinction between *da* and *de* becomes blurred at times (*de* is also used as a partitive as in *kvarono de* a quarter of) and may correspond simply to collocation (preferred expressions). For example, the distinction between the head noun *manko de* ('lack of') and quantifier *manko da* ('lack of') translates into different phraseologies in the corpus: *manko de* + sentiment: *manko de*

kompreno lack of understanding, *manko de kuraĝo* lack of courage, *manko de kontakto* lack of contact vs. *manko da* + material or verbal processes: *manko da vortoj* lack of words, *manko de vojoj* lack of routes, *manko da reago* lack of reaction.

2) Adjective phrases and adjectives in the noun phrase:

The cohesion of the noun phrase is indicated by the potential for agreement for every category in the NP apart from numerals and certain determiners. The correlative determiners (ending in *–iu*, *-ia*) agree for case and number as do adjectives: *tiuj gigantaj konstruaĵoj* 'those gigantic buildings'. As noted in the section on word order above, the plural as well as the accusative is also carried on into embedded structures, (*ili konsideris la landon* [*regantan la insulon*]: 'they considered the country [ruling-ACC the Island-ACC]'). Adjective placement is usually prior to the noun. In the corpus the Adj + N order occurs 14 159 times, compared with just 420 times for N + Adj. While some N + Adj forms are equivalent to Adj + N (*la tablo granda, la granda tablo* the large table), most N + Adj sequences have a clearly marked appositive or 'naming' function as in *amikino mia* a girlfriend of mine, *io nova* something new, *io mirinda* something amazing. The internal structure of adjective phrases is Specifier (*pli* more, *tre*, very) + Adjective + Complement (*ol* than). Some adjectives permit complement clauses: *konscia ke* aware that, *certa ke* certain that. The Adj + N structure is paralleled by Participle + N (*kreskanta trafiko*: rising traffic) but also by a less typical construction where the noun is preceded by a non-finite relative clause: *Tio estas la* [*penetranta ĉion*] *militismo* = 'that is the (penetrating everything) militarism' (that is the militarism which penetrates everything).

One particularly original aspect of adjectives in Esperanto involves the use of relative participle clauses which allow the placement of prepositional complements between the determiner and noun: D + PP + Adj (participle) + N. This is generally a Germanic formulation. The construction is analogous to the Participle + N structure mentioned above, and typically involves a passive participle in normal adjective position preceded by its associated prepositional complement: *la de vi solvita problemo* 'the problem solved by you' (the alternative form is however more usual: *la problemo solvita de vi-* the problem solved by you). The construction can also follow normal Participle + Agent order: *ĝiaj legantoj retrovu tuj la uzitajn de nia Majstro vortajn*

formojn 'so readers might find immediately the [used by our Master] word forms'. The corpus contains a handful of examples of this:...*por ke vi taŭge povu plenumi la <u>al vi donitajn taskojn</u>*: 'so that you may fulfill the tasks (that were) given to you'. This is also a Germanic formulation. Kalocsay mentions that a slightly different construction (D + ADJ + PP) is particular to Slavic languages which more readily admit additional modifiers between specifier and noun. In the corpus, not surprisingly, most examples come from the *Soveta Esperantisto* and at times the agentive preposition *far* is used: *la <u>ellaborita far la partio</u> sociala politiko* 'the (worked out by the party) welfare policy', *la <u>trudita far la imperiismo</u> vetarmado* 'the (imposed by imperialism) arms race' (as mentioned below, *far* is an invented preposition derived from *fari* 'to do'). This formula is extended to normal adjective + prepositional modifiers: (this time a Germanic formulation, but found in the *Soveta Esperantisto*) *la socialismo kreis por ŝi egalajn <u>kun la viro</u> eblecojn labori* 'socialism created for her [equal-with-men] possibilities to work' (i.e. 'created for her possibilities to work that were equal with men').

3) Numerals (*Fundamento,* Rule 4)

 Within the noun phrase numerals precede adjectives and follow determiners (they may also function as determiner or as head, in the absence of a head noun: *la du alvenis*: (they) both arrived). Number formation is consistent and resembles the combinatorial system used in radio communication: *dek unu* 'ten one, eleven', *sepdek* 'seven ten, seventy'. In addition, for some writers *unu* is variable, i.e. when used as both a pronoun and determiner (*unuj diris* 'some said', *unuj partioj* 'some parties'). The preposition *po* is used to resolve distributive expressions of the type 'by two', 'each' (thus *ni portos po du skatoloj* we'll carry two boxes each / at a time). The corpus contains uses with the accusative suggesting an unofficial adverbial use of this word may have emerged: **ni aperigis ĉijare la ĵurnalon po kvar fojojn:* we published this year the journal four times' (i.e. *po* = at the rate of).

4) Pronouns: (*Fundamento* Rule 5).

	Singular	Plural	Reflexive	Archaic
First Person	*mi*	*ni*	*mi / ni*	
Second Person	*vi*	*vi*	*vi*	*ci*
Third Person -impersonal -neutral -masculine -feminine	*oni* *ĝi* *li* *ŝi*	*ili*	*si*	

The use of pronouns in Esperanto is oriented to human gender as in English (*li* and *ŝi* can only refer to people and are not determined by grammatical gender), and to a basic distinction between person (perspective) and number (singular vs. plural). The first person is the same as for many European languages, whereas the second and third person systems are particularly close to English. The use of impersonal and reflexive pronouns and the role of zero-subject is however different to English. The following points address issues that are specific to Esperanto:

Ci - The pronoun *ci* was originally devised by Zamenhof to represent the archaic informal or Shakespearean 'thou' in translations and is used by some Esperantists as a familiar 'you' because of the influence of European languages (French *tu,* German *du,* Russian *ti* etc.). Zamenhof discouraged *ci* with the justification that different languages had different conventions for its use, and there may also have been an ideological motive to establish parity for all language speakers. There are no examples of *ci* in the corpus, although its use is attested among some reform-minded Esperantists. Conversely, a polite form *Vi* (i.e. capitalized *vi*) was widely used as a polite or plural second person form in early writing, and there are a number of instances in the literary sections of our corpus, including the early writings of Zamenhof.

Ĝi - The neutral *ĝi* 'it' is reserved for inanimates and non-human animates, although it has also been used as an impersonal or non-gender term for people (another

alternative is *tiu*: that one). Unlike English usage, children and pets are sometimes referred to as *ĝi* (as seen in recent editions of *Asterix* and *Tintin*). Although *ĝi* corresponds for the most part with the English use of pronominal 'it', Esperanto has an obligatory zero-subject for weather and time expressions (*pluvas* it is raining, *neĝas* it is snowing) and for statements where the subject is a general proposition rather than a specific referent (*dependas* it depends, *estas grave* it's important, *ŝajnas ke* it appears that…).

Mia, via etc. - Possessives are formed as adjectives (thus *mia*, my or mine, *nia* our or ours) although they are not used in the same distribution as adjectives (for example, they occupy determiner position in the NP and they cannot be preceded by specifiers like *tre* – very). When used as pronouns (i.e. replacing the noun phrase), possessives can be preceded by *la* and require no main noun in the phrase, as with numerals: *la miaj foriris* mine went away. Possessives are not used in reflexive contexts: *li levis la kapon* he raised his head, *li havis la poŝon plena* he had a full pocket.

Si – The reflexive pronoun *si* is designed to distinguish between third person reflexives and references to a third party (thus distinguishing between *ŝi amas <u>sian</u> aŭton*: she likes her car - her own car, and *ŝi amas <u>ŝian</u> aŭton*: she likes her car -some other person's car. This is a common distinction in Slavic languages, as in Russian:

'oni vidjat svoj dom' = *ili vidis sian domon* ('they$_i$ saw their $_i$ house')
'oni vidjat ihh dom'   = *ili vidis ilian domon* ('they$_i$ saw their$_j$ house').

The use of *si* is usually explained as 'reference back to the subject' and is also said not to extend beyond the clause. This causes hesitation in non-finite clauses and other modifying phrases. For example, in the sentence *li vokis sian kuzinon* [*loĝantan en <u>lia</u> domo*] 'he$_i$ phoned his$_i$ cousin [who lives in his$_i$ house]' the use of *sia* in the relative clause is avoided because it might be construed as referring to the cousin (<u>her</u>), despite the fact that *sia* is used unambiguously in the main clause to refer to <u>him</u>$_i$. Venture (1977) mentions a different example in Zamenhof's writing: *ĉiuj$_i$ dormis krom la direktisto$_j$ apud ?<u>sia</u>$_i$ direktilo* (all$_i$ were sleeping save the helmsman$_j$ at ?<u>their</u>$_i$ helm) where *sia* strictly speaking refers back to the subject 'everyone' not to the helmsman (to refer to the helmsman, <u>lia</u> must be used). With conjoined nouns, there is

also possible confusion: *Tiuj*ᵢ *estas la kuracisto Armin K. kaj sia*ᵢ *patro*. 'These (people) are Doctor Armin K. and his father' where strictly speaking *sia* refers back to *tiuj* (these) and make the reference illogical.

By convention, *si* is not used with first and second person pronouns, where reflexives are unambiguous (*mi lavis min* I washed myself). However, because *sia* is used as a derivational prefix (as noted under morphology, above), some relatively fixed derived forms such as *siatempe, siamaniere* break the convention when referring to non-subject elements of the sentence or first and second person subjects. The corpus provides numerous examples: ?<u>*Siatempe*</u> *mi havis ĉevaleton*: In my day I had a pony (not *miatempe*), *la gvidantoj, kunveninte en Bonn*, ?<u>*siamaniere*</u> *'celebranta' la 40an datrevenon* … the organizers …meeting in Bonn, in its own way 'celebrating' the 40th anniversary… (not *ĝiamaniere*). Rather than demonstrating inconsistency, these 'mistakes' show that modifying elements have their own internal sentence-like structures and apply constraints locally which are consistent with the rest of the system. In other words, the rule of thumb '*si* refers back to the subject' does not seem to cover cases where pronouns in modifying attributive phrases refer to the elements to which they are immediately attached.

Oni - The impersonal pronoun *oni* 'one' is part of Esperanto's voice system and coexists with the use of ergative forms for verbs, passive participles, passive verbs, the use of the accusative in OVS word order etc. As well as its more common use as a vague pronoun (where English tends to use plurals 'we', 'they' or non-specific 'you' and 'one'), *oni* is obligatory with di-transitive verbs (where there is an object and an indirect object) and where English might use a passive: <u>*oni donis al mi du librojn*</u>: 'Someone gave me two books, I was given two books'. The use of *oni* is always restricted to subject or (rarely) as a derivational prefix in words such as *onidiroj* 'rumors'.

3.4 Verb Phrases

Most aspects of the verb phrase in Esperanto belong to the discussion of morphology. The following examples from the corpus give a sample of the wide variety of possible constructions and set out the potential internal structure of all active verbs: ([ ] = obligatory element, ( ) = optional).

Verb forms

(Prefix)	[Verb]	(Voice)	(Suffix, Participle)	[Finite, Nonfinite]	
trans	*met*			*i*	'to transmit'
re	*met*	*iĝ*		*as*	'is re-placed'
	danc		*ad*	*is*	'went on dancing '
ek	*manĝ*			*os*	'will start to eat'
	manĝ	*ig*	*it*	*as*	'is force fed'
	vek		*int*	*is*	'has just woken'
	venk		*it*	*os*	'will be defeated'
	vol		*ant*	*us*	'would be wanting'

Participle forms

(Aux.)	(Prefix)	[Verb]	(Voice)	[Participle]	[Word class morpheme]
estas		*ŝanĝ*	*iĝ*	*ant*	*a* 'is changing'
esti	*trans*	*met*		*int*	*a* 'to have transmitted'
		kontent	*ig*	*ont*	*e* 'about to make happy'
estis		*elekt*		*at*	*a* 'was being elected'
estis		*vest*		*it*	*a* 'was dressed'
estas		*nask*		*ot*	*aj* 'are about to be born'

Participle forms are adjectival in that they agree in the plural and the accusative as attributes attached to a noun phrase or as attributive complements of the verb 'to be'

(*ili estis disŝirantaj ĉion*: 'they had torn everything up', they were having-torn-PL everything-ACC). Participles are also verbal in nature, since they also take complements in their own non-finite clause: '*la Statuto de UN* [*proklamanta rifuzon*]: the UN statute [proclaiming a refusal]'. In Esperanto predicative verbs can be formed as derived active verbs using participles (perfective, progressive) as part of the word root: (*manĝitas* gets eaten, *venkintos* will have vanquished).

Syntactically speaking there is only one 'voice' in Esperanto:

Active: *mi fermis la kovrilon* (I closed the cover)

Other voices can be seen as morphological variations of active sentences (i.e. all have morphologically active verb forms). Voice distinctions essentially signal semantic differences in the way that the verb relates to the subject:

Passive: *…kaj fermitis la pordoj* …and the doors were closed [compound form]
Middle: *li turnis sin al la strato* - he turned (himself) towards the street.
Ergative: *la trinkejo frue fermiĝis* - the bar (became) closed early.
Causative: *tiu ideo timigis min* - that idea frightened me (made me fear).

The equivalent participle form extends the active verb group to F + P (Finite + Predicate), allowing the finite verb *esti* to express tense while the participle expresses aspect:

Active: *li estas malferminta la pordon* - he has closed the door.
                 BE+PRES   ACTIVE PERF PART.
Passive: *malsama sinteno estas bezonata* - a different stance is needed.
                 BE+PRES   PASSIVE PROG PART.
Middle: *Consuelo klinis sin antaŭen* – Consuelo leant (herself) forwards.
                 PAST   SELF
Ergative: *la ŝtormo komenciĝis* – the storm had (become) started.
                 ERG+PAST
Causative: *mi entombigis la malĝustan katon*- I buried the wrong cat.
                 CAU+PAST

Apart from compound participles, further variations of verb use take the form of verb serialisation. As with copular verbs (*esti* to be, *ŝajni* to seem, *vidiĝi* to se been as) *ig*

(causative morpheme) and *iĝ* (ergative morpheme) are used serially where the attribute of the verb 'to be' is understood:

Sb	Vb	Ob	Ca
La eldiroj	*igis*	*min*	*silentema*	The words made me (be) silent.

Sb		Vb	Ca
La festivalaj renkontiĝoj		*iĝis*	*proskimaj*	The festival meetings got nearer (close).

As mentioned in the morphology section, only *igi* can also be used to create verb series where the object of *igi* is projected as subject of the new clause. Syntactically, the projected clause is seen to be embedded as complement of the verb:

Sb	Vb	Ob	[Vb	Ab]
Oni	*igis*	*min*	*stumbli*	*ĉirkaŭ cirklo.*

They made me stumble around a circle.

Verb complexes are typically formed by modal verbs in English and German, and although Esperanto has equivalents as in *ŝi povas demisii* – she can resign, these verbs can be considered for the most part as normal transitive verbs with verb phrase complements. Modal verbs play a role in the formation of verbal meaning, usually in terms of modality (verbs such as *povas* can*, eblas* 'be able to'*, devas* must) and in terms of aspect (*daŭras* to continue, *resti* to remain, to stay). Serial verbs tend to allow the grammatical subject to be carried on from the original clause (a longer example of this is: *Se vi bonvolos / helpi / prizorgi la infanan kongreson...* 'If you want to please / help / look after the children's congress…') In Esperanto modal verbs take infinitive complements:

Sb	Vb	[Vb]	
Mi	*volis*	[*telefoni*]	He wanted to telephone.
Mi	*devas*	[*fari ĉion eblan*]	I must do everything possible.

Other aspects of the verb phrase involve specifiers, grammatical adverbs that determine the temporal scope, the polarity and sometimes the modality of the verb.

Specifiers have a similar role as determiners in NPs and are usually placed as near as possible to the left of the active verb. Time-specifiers are to be found in close proximity to the active verb: *jam* 'already', *ĵus* 'just', *nun* 'now', *baldaŭ* 'soon'. *Jam* in particular is used to shift temporal aspect to a perfective sense, and is regularly used in conjunction with the present tense in the complex adverb *jam de longe*: *li jam de longe malsanas* 'he has been ill for some time'. Similarly, *ĵus* shifts the aspect of the active verb to a progressive sense (*ŝi ĵus traktas ĝin* she is just dealing with it, *mi ĵus ricevis jenan noton* I've just received this note, i.e. the event is completed but the result is still ongoing). As adverbs these can be placed elsewhere in the sentence, often initially. The placement of *ne* is more constrained, and as a logical operator its use elsewhere in the sentence is likely to be a sentence negative (and thus used as a tag: *ne, estas malĝuste*: no, that's wrong, *ĉu ne?* isn't it?). Although regularly placed to the left of the active verb, *ne* is rather markedly placed to the right when used with identifying complements and thus functions a modifier of the noun phrase (*tio estas ne Alfred White*: that is [not Alfred White] (with the implication that it is someone else). *Ne* as modifier of noun can thus be contrasted with determiners of the noun *neniu, nenia*. *Neniel* 'in no way' is at times used in specifier position close the verb, and this may be seen as an emphatic expression and an analytic alternative to *ne*.

3.5 Prepositional Phrases

A number of Zamenhof's rules involve prepositions in the *Fundamento*. Rule 8 states that prepositions have one fixed meaning, although it is clear that this is difficult to uphold this in all cases (for instance *al* means 'to, towards' but is also used in a more abstract sense for indirect objects, such as *prepari la junularon al familia vivo* prepare the young generation for family life). Other reform projects had developed highly sophisticated propositional sets in the same logic that led Esperanto to attempt to reduce homonymy in the vocabulary and Zamenhof had to resist attempts to divide *de* for example into different paronyms (*de, di, da, du* etc.) to represent the different functions of this preposition. Nevertheless in the basic design of the language there was an attempt to delimit usage while keeping basic forms similar to Latin originals. The preposition *de* still has a large number of functions (meaning variously 'of, from, off', as well as being used as an agentive after the passive and as a preposition

signaling the possessive as in *la libro de Petro* Peter's book). However, *de* can still be contrasted with *el* (reserved for the specific direction 'out' as well as for signaling materials: *el ligno* 'made out of wood') as well as *da* ('quantity of': used in determiner expressions). *Por* ('for', used to signify 'for the benefit of' as well as a complementizer meaning 'in order to') may be contrasted with *pro* (because of), *pri* (about, regarding) and *per* (by, by means of).

Prepositions in Esperanto are sometimes used redundantly, often in conjunction with prepositional prefixes attached to verbs (*eniri en* to enter into) or at other times, complex prepositions can be formed (*el sub*: from under). The following example demonstrates some of these uses:

Se mi estos suprenlevita <u>de sur</u> la tero, mi <u>altiros</u> <u>al</u> mi ĉiujn homojn.
If I am (upwards)-carried (<u>from on</u>) the earth, I shall (<u>to</u>)-pull (<u>to</u> me) all Men.

'If I am to be carried up from the Earth, I shall attract all Men to me.'

In the *Fundamento,* Rule 8 states that there are two cases in Esperanto: nominative (subject) and accusative (object) while the prepositions were designed to replace the cases of other languages. A trace of the genitive persists however in the correlative series (*kies, ties*: whose, this one's etc.). In syntactic terms, prepositions are involved in three basic sentence functions: adjuncts, indirect complements, nominal modifiers:

Adjuncts: The preposition and its phrase have an adverbial function, modify the verb or the clause and can often take any position in the sentence: <u>*En la mezo, troviĝis la trezoro*</u>: 'in the middle was (found) the treasure', *ili vojaĝis <u>en la ĝangalo</u>* 'they traveled in the jungle'. Rule 13 indicates that direction is expressed by the accusative *–n*. A similar function of the accusative exists in a number of languages and this appears to be only used in adjuncts: *ili vojaĝis <u>en la ĝangalon</u>* (they traveled <u>into</u> the jungle).

Indirect complements: The preposition and its phrase are complements of the verb (i.e. are determined by the choice of verb) and are somewhat more fixed in position (*tio kondukas <u>al malkresko</u>* that leads to a decline, *ŝi donacis ĝin <u>al ili</u>* she had donated it to them, *tiu problemo dependas <u>de la registaro</u>* that problem depends on the

government. These are essentially collocational properties of verbs, and are similar to case marking after verbs in Germanic and Slavic languages, as well as Latin and French. One property of an indirect complement preposition is that it can not be used as prefix to a verb (compare from the examples above: *alkonduki* = complement 'to lead to', <u>entroviĝi</u>= modifier 'to be found within', *aldonaci* = complement 'to donate to'*, <u>prizorgi</u>* = modifier 'to care about'). It is noticebale that *al* and *de* can only form verbal complement prefixes when they are used as modifiers rather than complements, that is their interpretation is likely to be limited to parallel compound constructions (*aliro* = approach 'to-going', *deveno* origin 'from-come': as mentioned in section 2.5).

Nominal qualifiers: Prepositions are also used within the noun phrase as complements or modifiers of the noun. These generally follow the main noun: *manifestiĝo de naciismo* a manifestation of nationalism (complement), *junulo el Albanio* a youth from Albania (modifier). As in other languages, where the noun is derived from a transitive verb, there may be ambiguity if *de* is used as the linking preposition (*de* also implies agency) and so *malamo de la faŝistoj* is conventionally interpreted as 'hatred by Fascists' whereas *malamo al la faŝistoj* conveys 'hatred (directed towards) the Fascists'. In most cases, no ambiguity is possible and the 'normal' nominal construction is *la riparo de pontoj* 'the repairing of bridges'. The *al* remedy is however sometimes used in the corpus where no ambiguity is in fact possible: *la serĉado al laboro*: 'the search for work' and this may instead be interpreted as an emphatic paraphrase (?the search towards a job = > the long search for a job). It is also possible for *je* to be used instead of *al* to indicate a complement such as this. As noted below, this role would be consistent with other uses of *je*, but is not frequently exploited.

The use of prepositions in Esperanto has perhaps been one area of the grammar that has been the most subject to change and innovation. There are three areas of particular interest: adverbial prepositions, agentive markers and the neutral proposition *je*.

1) Adverbial prepositions (ADV+ *de*). There have been a number of innovations which have affected, and in some cases reduced the scope of prepositions. Of particular importance are adverbial constructions which tend to replace or operate in

parallel with prepositional phrases. One innovative and very common construction involves redundant, or at least 'emphatic' prepositions formed by <u>adverb + *de*</u> (*ene de*: within, *pere de* by means of, *sube de* underneath), which replace the simple use of the preposition (*en* in*, per* by*, sub* under). By analogy with this process, lexical items are systematically reformulated as adverb-prepositions and this proves to be highly productive. The corpus contains 491 examples of adverb-prepositions involving over 20 forms, the most frequent being: *fare de* by the action of (discussed below), *fine de* at the end of, *surbaze de*: on the basis of, *escepte de* except, *kadre de* in the framework of, *komence de* at the beginning of, *meze de* in the middle of, as well as numerous uses with dates (*novembre de 1984* in November 1984). The widespread use of this form is likely to affect the distribution of 'traditional' prepositional phrases of the type *laŭ la opinio de* (according to the opinion of) or *dum la vizito de* (during the visit of) which can be systematically reformulated by *laŭopinie de* and *vizite de*. Critics of this tendency have pointed out that the compound form may at times obscure the relation signaled by a preposition, although many European languages use prepositions idiomatically in full prepositional phrases (in English, *at, in, on* etc. are often used for the same function). The productive use of lexical root + [*-e de*] to replace these forms can be viewed as a considerable streamlining of otherwise idiomatic structures. Janton (1994:70) has also noted a parallel tendency to replace prepositions by (modifying) prefixes in verb compounds as in *kapjesi* (*jesi per kapo*= 'to say yes by the head', to nod), *luktakiri ion* (*akiri ion per lukto*= to acquire something by struggle) as well as adjectival compounds (*la trafiko per tera vojo -> tervoja trafiko*: ground traffic (traffic via the ground).

2) Agentive markers *de- / per- / far-* Another particularly widespread innovation that emerged early on in Esperanto's development involved the use of prepositions other than *de* after passive participles. *De* unambiguously signals the grammatical subject of any passivized transitive verb (*La frazo estis interrompita de subita ekbrilego*: 'The phrase was interrupted by a sudden flash'). However, it has been argued that the subject is not always the agent, and in the following examples, *de* introduces either the medium or the instrument of the action rather than the true agent: *mia animo estis kaptita de malgajo* 'my soul was captured by unhappiness', *ekzistas danĝera insulo plenigita de murdantaj bestoj* 'there exists a dangerous island full of (filled by) murderous beasts' (both instances of *de* here are the original forms used in

the corpus). In many cases *per* (by, with, by means of) has come to be used in this context to signal the instrument of any action (*karakterizita per alta klereco*: characterized by high intelligence, *pravigata per rezonado*: a fact proven through reasoning). And *far* (derived from the verb *fari* to do' and perhaps by analogy with French *par*) was proposed to unambiguously signal a willing agent *(akceptita far loka grupo* 'accepted by a local group, *proklamita far la UN Jaro de Paco* proclaimed by the United Nations Year of Peace). In the corpus past participle + *de* is still the most frequently used form (297) for all of these uses. *Per* is used in 94 instances, while *far* is used 14 times. However *far* has found wider application with nominals (42 instances: *ekspluatado de homo far homo* exploitation of humans by humans), and *far* may also be expressed as an adverb-preposition *fare de* (34 instances): *senarbigo en Tibeto fare de la ĉina registaro*: 'deforestation in Tibet by (the action of) the Chinese government', *pro miskompreno fare de la konsilio de Eŭropo* 'because of a misunderstanding (generated, caused) by the council of Europe'. As with many developments of this kind, *fare de* can be attributed specifically to an individual, in this case the French writer E. Grosjean-Maupin (Piron 1989a:140).

3) The neutral preposition *Je* - Rule 13 of the *Fundamento* stated that the preposition *je* exists for ambiguous expressions, but is replaceable by *–n*. Thus when there is no clear semantic reason to justify the use of one preposition, the neutral form *je* is used (for example in complements of adjectives: *plena je ĝojo*: full of joy or to avoid ambiguity in idiomatic expressions *mi vetas je ĉevaloj* 'I bet on horses'). *Je* is a relatively frequent preposition and it is used in a much wider set of contexts than Zamenhof envisaged (Wackrill (1907) lists six contexts in Zamenhof's early writing, all post-verbal except for *plena je* and *je la flanko* 'on the side'). The main uses of *je* in our corpus involve expressions of date and time (*je la unua (horo)* at one (o'clock), *je la komenco* at the beginning, *je la fino* at the end, *je tiu tempo* at that time, *je novembro*, in November), spatial expressions where the noun carries the main meaning (*je la pinto* at the summit, *je unu flanko* at / on one side), prices (*je kosto de* at a cost of, *je du mil spesoj* at a cost of 2000 spesos). After most verbs *je* has a spatial sense (usually translated by <u>at</u> in English). *Je* is only used as an indirect complement after certain verbs (*kredi je* to believe in, *fidi je* to trust in, *minaci je* to threaten with). In noun phrases *je* has a number of complement uses although they are rather infrequent (*rajto je laboro* right to work, *pretendantoj je la rolo de decidantoj*

pretenders to the role of decision-makers, *rekompenco por via fido je li* a reward for your faith in him) as well as other uses as modifier (*subvencio je du fontoj* a grant from two sources, *braveco je patriota servado* courage in the service of one's country, *sen diferenco je raso, sekso, lingvo kaj religio* without distinction in race, gender, language and religion).

Rule 13 states that the accusative may be used to replace *je*. For dates and times, the accusative is consistently used more frequently than the *je* form. The accusative can even be seen after intransitive verbs, where an indirect complement is therefore assumed (Zamenhof cites the following possibility: *mi kontentiĝas je tio* I am happy with that -> *mi kontentiĝas tion* 1962:67). Although the accusative does usually replace adjunct / modifier uses of *je* (*je la alia flanko = la alian flankon* on the other side) it is not used to replace indirect complement expressions (hence the unacceptability of **mi vetas ĉevalojn* I bet *horses). Zamenhof pointed out that accusative replacement was also impossible in the presence of two complements: *pardoni malamikon* to pardon an enemy, *pardoni kulpon* to pardon a fault, but not **pardoni malamikon kulpon* (instead: *pardoni al malamiko kulpon*: to pardon a fault of one's enemy). Since indirect complements are largely dependent on the choice of verb, *je* often appears to occur in idiomatic contexts. The nominal use of *je* can also be expressed by an accusative: *la kredo je diabletoj-> la kredo diabletojn* (the disbelief in demons), although this use is not attested in the corpus. Janton (1994:73) however attests the following possibilities: *enkonduko Esperanton* 'the introduction to Esperanto', *inda atenton*: worthy of attention and the derived adverb: *rilate aferon* relating to business. While Zamenhof sanctioned the use of these expressions, they have not been widely used, and of all these forms only *rilate*+ACC occurs widely in the corpus (with 29 occurrences it appears to be slightly more frequent than *rilate al* with 25). The potential to replace *je* by the accusative has nevertheless extended the theoretic use of the accusative to other areas. This makes it difficult to distinguish between accusative replacements of adjuncts (where the preposition is not obligatory) and indirect complements replaced by the accusative (*mi ne devios eĉ unu paŝon*: 'I shall not err (by?) even one step', *tiu atento atestas tion, kio estis la plej terura*: 'that expectation testifies (to?) that-ACC which is the most terrifying').

Overall, the use of the accusative and *je* can be seen as part of a generalized process by which Esperanto erodes the use of prepositions in areas where they are idiomatic or ambiguous in a number of languages. But prepositional phrases, the accusative and *je* are in turn prone to replacement by adverbs and the formation of compounds where the modifier or complement may be used as a prefix. Thus *manpremi* can be read as *premi per la mano* 'squeeze / shake by the hand' (or perhaps *premi la manon* to squeeze the hand). And as noted above (in the grammatical morphology section 2.3), the creation of derived adverbs eventually affects adjunct expressions which are usually reserved for *je*:

> *je la lasta fojo-> la lastan fojon -> last(a)foje* the last time,
> *je la pasinta fojo-> la pasintan semajnon-> pasintsemajne* last week.

Given the tendency to replace prepositions and accusatives by adverbs a general shift pattern can be posited:

> Prepositional phrase
> → accusative (implying the potential use of *je*)
> → Adverbial modifier (by prefixing, parallel derivation or by creation of an -*e+de* preposition).

As with much reformulation in language, these processes are not permanent and they are certainly not applied to all phrases in the language. For example, compound words formed by the lexical morphemes appear to be immune to the addition of adverb morphemes. The process demonstrates however that general principles of morphological change have begun to impinge upon the syntactic properties of the language and that the language is in a healthy state of evolutionary flux.

4 PHRASEOLOGY

Phraseology is the umbrella term for a variety of language phenomena, and is one of the least well-defined areas of linguistics. In this section we set out the basic principles of phraseology and argue that Esperanto displays the same conventional properties as other natural languages. Generally, a speaker's knowledge of phraseology involves three areas of communicative competence: lexical paradigms (the available resources of vocabulary open to the speaker), idiom structure (the conventions of collocation, formulaic and fixed expression) and cultural reference (the social references that identify the insider and delimit membership of the community).

4.1 Lexical paradigms

There are three lexical systems in Esperanto: (1) the basic 'official' vocabulary of the *Fundamento*, (2) constructed words (innovations derived from rule 11 – 'compound words are formed by a junction of roots, the head word at the end') and (3) borrowed words (neologisms permitted by rule 15 – 'the lexicon may be built from any internationally recognizable words'). One of the monolingual dictionaries, the *Plena Vortaro* (the 'complete dictionary') indicates four vocabulary types: words belonging to the *Fundamento* (marked *), officialized words (marked with capitals), unofficial neologisms derived from Zamenhof's writings (marked z) and technical terms (marked with a thematic symbol).

Esperantists are familiar with the dual nature of compounds and neologisms, and debates about which forms are more 'official' or 'logical' are part of the Esperantist culture. Normal processes of neologism and borrowing account for the majority of Esperanto's current vocabulary. Piron (1989a) cites a number of ordinary lexical items which were never part of Zamenhof's original lexicon: *teko* (a case, as in briefcase), *novelo* (a short story), *oferto* (an offer). Some basic roots have been modified by the community for simplicity or euphony. Thus *aŭtenta* is used alongside *aŭtentika* (authentic) and *spontana* seems to have replaced *spontanea* ('spontaneous') (Piron 1989a:133). However, this gradual change is perhaps not as widespread as change through morphological analogy. In particular, the agglutinating structure of the

language has led to a number of competing lexical sets, often opposing words derived from donor languages against regular constructions derived from the normal rules of morphology. Piron cites the development of the word *civilizo* instead of the official *civilizacio* working backwards from the common verbal form *civilizita* (civilized). The words *civilizo* and *civilizacio* are both used in the same context despite their differing origins and status, and our corpus shows that they have roughly similar frequencies of use: 17 and 14 respectively. Other functional synonyms include *situo* vs. *situacio* ('situation', in the corpus *situacio* predominates over *situo*) and *poluo* vs. *polucio* ('pollution', in this case the corpus shows that *poluo* has replaced *polucio*).

Even using basic roots and the normal rules of derivation, synonymous words can be created, a fact that is often forgotten in descriptions of the language as 'logical' or 'synonym-free'. We have mentioned in the section on morphology the dual system in place for countries such as *Brituljo* and *Britio* (Britain) but the process extends to other compounds formed by morphemes, as in the case of *policano* 'member of the police' which coexists with *policisto* 'police officer' (both have roughly equal representation in the corpus). Similarly: *vidigi* to make seen / *montri* to show, *manĝigi* to feed / *nutri* to feed. Although these terms appear to coexist with little comment, others appear to be more controversial, possibly because of the technical or institutional implications at stake, and in the case of borrowed neologisms there is at times hesitation over possible alternative forms. For example, for some time several terms were used for 'computer': *komputoro, komputero* (both single root words) which were eventually supplanted by *komputilo* 'computing tool' (root + morpheme). In some cases, the borrowed neologism displaces the original compound form (*trajno* 'train' appears to have replaced *vagonaro* 'train, set of wagons', which is now felt to be archaic and possibly equivalent to 'railroad' in English).

In most cases, terms acquire different meanings. Semantic displacement can be seen in the case of such borrowed words as *rendevuo* 'rendez-vous' and *mitingo* 'political mass meeting' or 'rally'. Both words have more general meanings in their donor languages ('rendez-vous' from French and 'meeting' from English) but now coexist with the original and surviving Esperanto terms for 'encounter' *renkontiĝo* and 'meeting' *kunveno*. As in other languages, the increasing use of the term *komputilo* (computer) has similarly led to a more specific use of the original word *komputi* (to

compute). Not only do lexical items and neologisms have different semantic roles, but they are also likely to develop different grammatical functions. For example, we observe from the corpus that few unofficial or derived verbs introduce complement clauses (as with *deziras ke* 'desired that'*, konkludis ke* 'concluded that') and no derived words are used to introduce noun complement clauses (as with *la fakto ke* the fact that, *la ideo ke* the idea that). While there is no theoretical reason why certain neologisms or compound forms cannot take certain grammatical constructions, in practice they are not exploited in the same way. It could be argued that certain syntactic features have been 'borrowed' at the same time as the vocabulary that motivates them. This is a fundamental property of phraseology (as we see in the section on collocation, below) and is consistent with processes of borrowing and neologism in other languages.

Many Esperantists are uncomfortable with the idea of variation and near-synonymy in the vocabulary of the language, but as Janton (1994) has pointed out multiple vocabularies are an integral part of Esperanto's system of register and style. Variant words should be considered as mutually exclusive, but as belonging to different repertoires available to the speaker. Neologisms (especially single root words) are associated with a learned, literary or journalistic style and for immediate recognition they often depend on wide reading or knowledge of the donor language. Where society or technology moves on, some compound terms either fall into disuse or are reassigned. As with *vagonaro* above, the compound word *malsanulejo* was eventually replaced by *hospitalo*, while *malsanulejo* is now somewhat archaic and means roughly 'sick-house'. On the other hand, some compounds dating from Zamenhof's original texts remain as part of the core vocabulary despite the fact that they bear little resemblance to international forms. Wells (1989:58) cites Zamenhof's preference for the derived form *lernejo* ('learning place') as opposed to a possible international form derived from forms such as *school, école, Schule, escuela* etc. Interlingua, Occidental and the other reform projects which followed Esperanto attempted to capture these international forms, although as noted in our section on morphology above, this often resulted in the abandonment of the monomorphemic principle and of any attempts to provide these languages with a creative derivational system for word-building. Despite the objections of reformers, it could be argued that compounds such as *lernejo* are in the long-run preferable to established western terms, and it has often

been noted that Chinese and Japanese speakers often prefer these analytic or motivated compounds over Western-style synthetic forms (Hagège 1985). The concept of gravity for example can be expressed in Esperanto by *teraltiro* (Earth - towards - pull) or 'Earthwards-attraction' (a similar expression exists in Chinese), although Esperanto also uses the form *pezoforto* (weight - strength). In addition, compound forms may prove popular because they are 'indigenous' to Esperanto and their use in spontaneous writing and speech plays an important part in Esperanto word-play. Some forms are made deliberately long for humorous effect (the corpus gives us *bonfartigilmalaĉetloko* 'the place not to go to buy products that make you feel better' (well-fare-CAU-TOOL-OPPOSITE-buy-place-N). Compound homonyms are also an important source of material for puns; one often-cited example being *sentema*: 'themeless' (*sen – tem - a* without a theme) or 'sensitive' (*sent – em - a* tending to have feelings). In addition, compound forms are often formed spontaneously when the speaker has forgotten the original word, and the general tendency to exploit the derivational morphemes is often observed in the variety of possible synonymous forms (*adiciilo* addition-tool*, sumilo* sum-tool*, nombrilo* number tool, for 'calculator') and the formation of one-off redundant forms (**medikaĵo* 'medicine', **teatrejo* 'theatre(house)' (examples attested by Piron 1989a:136).

Janton (1994:82) claims that some compound forms have come to replace original or official terms, especially in literary writing. A search for Janton's examples in the corpus bears this out in some cases: the compound *vinbero* ('grape') (7) has come to replace *uvo* (no occurrences), while *elpensi* ('to invent, to think out') is on an equal standing with *inventi* (both with 25 occurrences). Compound terms can sometimes be considered clumsy or old-fashioned and in an opposite tendency to the one described by Janton, many have been replaced by single-root neologisms. Wells (1989) states that he would replace the word for 'left' *maldekstra* by *'liva'* while Zamenhof and other early writers often preferred to replace compounds by neologisms to vary their expression for literary purposes. This also appears to be the usual direction of borrowing in technical dictionaries, although there is less evidence of this in our corpus.

Kalocsay (1963: 54) cites a number of examples of existing compounds for which 'literary' replacements have been proposed. Their distribution of use can be gauged

from our corpus (all figures include counts of inflected and derived forms, excepting *mal-* derivations such as *malsupreniri*):

Neologism	Compound form	
Abismo (18)	*Profundaĵo (4)*	(abyss)
Angoro (7)	*Korpremiteco (0)*	(anguish)
Ascendi (4)	*Supreniri (41)*	(to ascend)
Aŭroro (7)	*Tagiĝo (4)*	(dawn)
Ferio (21)	*Libertempo (10)*	(holiday)
Festono (10)	*Florplektaĵo (2)*	(garland)
Giganto (25)	*Grandegulo (2)*	(giant)
Kalva (3)	*Senhara (2)*	(bald)
Karambolo (0)	*Kunpuŝiĝo (6)*	(collision)
Sino (5)	*Tronkantaŭo (0)*	(chest)

From this it appears that neologisms are generally more widespread than the original compound forms, and the fact that a relatively small corpus includes a sizable number of these words is significant. The relative rarity of most of the compound forms can be explained in some cases by their complexity and the fact that it is at times difficult to interpret them precisely (*grandegulo* can simply mean 'great big guy'). For some speakers, however the compounds may still tend to be more transparent (*tagiĝo* 'becoming of the day' is perhaps more accessible than *aŭroro*). On the other hand, those who do recognize the neologisms often find the words more evocative (this is certainly the case for *abismo* vs. *profundaĵo*). Some forms are particularly close to the core of Esperanto and are not that easy to displace even though the new word is arguably more meaningful for some speakers. Thus *supreniri* is more frequently used than *ascendi* but also has more morphological derivations and applications. If some single root neologisms tend to replace all but the most central lexical compounds, the case of *mal-* (the prefix 'opposite', discussed above in the section on morphology) appears to contradict this pattern. Kalocsay (1963) compares several *mal-* forms with their proposed counterparts, and here we have observed their frequency of use in the corpus (including some examples cited by Janton):

Neologism	Compound form	
Aflikti (4)	*Malĝojigi (3)*	(to depress)
Aperti (18)	*Malfermi (135)*	(to open)
Disipi (8)	*Malŝpari (10)*	(to waste)
Humida (1)	*Malseka (25)*	(humid)
Liva (0)	*Maldekstra (20)*	(left)
Neglekti (2)	*Malzorgi (1)*	(to neglect)
Noci (0)	*Malutili (12)*	(to harm)
Olda (3)	*Maljuna (61)*	(old)

The difference in distribution between these derived compounds and the previous list is striking. Even though the neologisms in both lists generally involve more specific, literary meanings than their compound counterparts, half of the *mal-* forms in the second list are significantly more frequent and generally have a wider set of applications than their proposed replacements. This suggests that many *mal-* words are more fully integrated in the language. Kalocsay claimed that *mal*-derived words were lexical items in themselves, and this may serve to explain the distribution of these forms compared with the compound forms in the first list. In section 2.2 (lexical morphology), we discussed the specific nature of derivational morphology in Esperanto and suggested that the system was essentially divided between lexical items and lexical morphemes. Whereas in the first list we are dealing with words whose scope is affected by various lexical morphemes, in lexical derivation with *mal-* a radically different word appears to be created (this may also happen with suffixes indicating people such as *ulo* and perhaps also the participles which conventionally signal people *-anto*, *-ito* etc.). To conclude, Esperanto appears to fully exploit both systems: morphological derivational is widespread but does not preclude neologisms (as in the first list) while some forms of morphological derivation happen to be more fundamental than others (as in the second list).

4.2 Idiom structure

Idiom refers to relatively fixed expressions that often have a pragmatic or rhetorical function extending beyond their literal meaning (Fernando 1996). While the idiomatic nature of language has become established in mainstream linguistics (Sinclair 1991, Halliday 1994), the study of idioms and other fixed expressions has been slow to

develop in the study of Esperanto, largely because of the lack of large text archives and the tools to exploit them, but also because discussion has been limited to marginal expressions and to proverbs. Some research has been carried out on the idiomatic nature of Esperanto's derivational morphology, in particular in relation to the concept of 'compositionaltiy' (the extent to which the interpretation of a given compound is predictable or simply down to convention: Schubert 1993, Dashgupta 1993). Other work concentrates on 'phraseological units', usually expressions which have come to be culturally recognized as part of a speaker's knowledge of the language (Fielder 1999). In this book, we claim that idiomatic expression is much more widespread than simply the use of marked clichés and proverbs, and that all phrases essentially take one of two forms in Esperanto: metaphor (the largely deliberate use of non-literal expressions and idioms) and collocation (the largely unconscious or predictable use of fixed expressions which permeate everyday language). Our discussion differs somewhat from the usual study of semantic compositionality: we see idioms as unusual expressions which are recognized as such in the speech community, while collocations are 'usual way of saying things', fixed expressions which have to be known to speak the language fluently (Gledhill 1999).

1) Metaphor

Many expressions are used non-literally in Esperanto. For example, *forpasi* 'to pass away' = to die, *zumi* 'to buzz' = to potter about, *celi* 'to aim' = to get at / mean something. Some expressions, especially compounds, involve a non-literal sense derived from the donor languages (for example, *librotenado* from English 'book-keeping') or have emerged because of generally expressed euphemisms (*necesejo* 'the necessary place': the toilet). Cruse (1984) has argued that this process leads to permanent semantic shift, recognizable for example in the virtually exclusive use of *ĉef-* (originally 'chieftain, leader') as a prefix meaning 'main, leading'.

Proverbs and clichés are complex and very fixed forms of metaphor. They generally involve truth-statements and have an element of word play about them. In the *Fundamento* Zamenhof equipped the language with a large number of proverbs, often attempting some form of humorous rhyme: *ne ŝovu la nazon en fremdan vazon*: 'don't shove your nose into a foreign vase' = mind you own business, *granda frakaso en*

malgranda glaso 'a big crash in a small glass' = a storm in a teacup*, unua atesto estas la vesto* 'a first clue is the dress' = the clothes make the man. In addition, a number of proverbs have been imported into the language from various sources, although very often the original expressions themselves are often difficult to trace. From Yiddish: *juĝantoj devas havi du orelojn* 'judges should have two ears' = there are two sides to every story, Polish: *vivi kiel koko en butero* 'to live like a cock in butter' = to live in the lap of luxury, French: *dika legomo* 'fat vegetable' = fat cat, English: *vivi hundan vivon* = lead a dog's life, Hungarian: *la pesto formanĝu ĝin* 'let the plague eat it' = to hell with it, etc.

Although most proverbs are rarely invoked in the general language, they form part of the basic repertoire of metaphors which the speaker may allude to (*fera mano* 'an iron hand', *amata ĉevaleto* 'hobby horse'). By writing down a set of proverbs, Zamenhof effectively created an oral history of the language, a corpus of expressions to dip into and cite or reformulate. Esperantists for example use the word *muŝo* in a series of contexts which can ultimately be derived from Zamenhof's expression *havi muŝon en la kapo* 'have a fly in the head' = to have a bee in the bonnet. However, few of the original expressions listed by Zamenhof (cited in Butler 1933) can be found in the corpus. A number of Zamenhof's shorter expressions involving adverbs are still widely known as basic idioms in the language (*dirite, farite* no sooner said than done, *vole-nevole* willy-nilly, *rapide senvide* more haste less speed, *la afero iras glate* everything is going smoothly). Idioms are notoriously difficult to find even in large corpora (Fernando 1996) largely because they are idiosyncratic and often simply transmitted from one dictionary to the next (few Esperantists will be familiar with *ĉe tro ĝentila ekstero mankas sincero*: too nice an exterior hides sincerity). The general lack of idioms and proverbs in the corpus may also be due to the fact that they are often reformulated or partially quoted for rhetorical effect. Short extracts of proverbs and clichés are often used in film and book titles (for example in English many James Bond titles are reformulated idioms or proverbs) and Esperanto similarly enjoys a growing source of fixed expressions from various works which are commonly known and can be cited in part (*mistero minora* a minor mystery, *infana raso* the child race, *peramuza gerilo* a 'fun' guerrilla war). Jokes and puns abound in titles. For example the expression *vole-nevole* (willy-nilly) is reformulated in the novel *Vole-Novele*

(willingly novel-ly) by R. Schwartz, while C. Rossetti's *kredu min sinjorino* 'believe me, my dear' becomes *knedu min sinjorino* knead me, my dear.

2) Collocation

Collocation involves the tendency for words to be selected in more or less strict combination. Collocations are more widespread than proverbs or canonical idioms of the type 'it is raining cats and dogs'. Many collocations obey the normal rules of grammar and simply represent a preferred way of saying things: *vera miraklo* a real miracle, *paco kaj amikeco* peace and friendship, *krokodilaj larmoj* crocodile tears, *partopreni* (= *preni parton*) to take part, to participate. 'Non-canonical' collocations obey different grammatical principles to the norm or involve unique formulations *dank'al* 'thanks to', *plenplena* 'full-full' = completely full, *finfine*: 'end-end-ADV' = at last, *ju pli granda ...des pli bona* (the bigger… the better), *post kiam* ('after when' for the conjunction 'after' instead of the potential **post ke*) and *antaŭ ol* ('before than' for the conjunction 'before' instead of the potential **antaŭ ke, *antaŭ kiam*). Although seemingly inconsistent, these are all frequent expressions and are taught as such in the basic teaching materials of the language. Non-canonical collocations are often more frequently used than their canonical equivalents, and may even motivate other forms by analogy. For example, the expression *plena de* 'full of' can be seen as non-canonical in origin, since other complement constructions after adjectives use the neutral preposition *je* (*riĉa je* rich with, *malsana je* ill with, *inda je* worthy of). *Plena de* is so frequent however, that this phraseology has come to infect other forms, for example: *libera je* coexists with *libera de* 'free from'. Lexical compounds which obey Kalocsay's principle of parallel compound formation (see lexical morphology, above) can also be considered to generate non-canonical collocations (i.e. prefixes which are complements of the main element): *voĉdoni* = to vote 'to give voice', *grandanima* = 'broad minded, magnanimous' (but not *martelbati* - to hit with a hammer, nor *plenigi* 'to make full' which are both normal compounds formed by modifiers and predicative adjectives of another element in the sentence). In addition, 'grammatical collocation' (also known as colligation) occurs when a limited set of lexical items collocate with specific syntactic features. This includes prepositional verbs (*dependi de* depend on, *kredi je* to believe in), complement clauses (*la ideo ke* the idea that, *gravas ke* it is important that etc.) and complement adjectives (*interesa ke* interesting that, *sufiĉa ke*

sufficient that). As a final classification, formulaic expressions include idiomatic, conventionalized collocations learnt as whole units. They often have fixed rhetorical uses: *k.t.p.* or *kaj tiel plu* 'et cetera', 'and so on', *pli kaj pli* more and more, *ne nur...sed ankaŭ* (not only… but also). It is often possible to predict most of a formulaic expression simply by referring to part of it. This leads to abbreviated forms for politeness formulae, and Esperanto uses a number of these curtailed expressions: *kiel vi*? (how are you? = *kiel vi fartas, kiel vi statas*), *kiomas*? ('what time is it?' = *kioma horo estas*), *ĝis!* (good bye =*ĝis la revido* 'until the next sight').

Without reliable data, the extent to which collocations are consistently used by speakers is difficult to judge. For example, one can not judge intuitively whether in English one should say 'to take a decision' or 'to make a decision'. Instead, a computer-based text corpus can be used to establish whether a collocational pattern has become established across a large series of contexts and to observe patterns of usage which may sometimes go against subjective intuition or received wisdom for any particular distinction. To give an example in Esperanto *danki* 'to thank' can either be followed by the preposition *pro* 'because of' (*dankon pro la averto* thanks for the warning) or by *por* (*mi dankas vin por la oferto* I thank you for the offer). Zamenhof used *por* (for) consistently in his early writing, although he maintained that *pro* (because of, due to) was a possible alternative. In this case, the corpus contains uses of both, although there are twice as many occurrences of *pro* (28) as of *por* (17). This kind of ambivalent evidence is a useful corrective to those commentators who attempt to claim that usage is wholly consistent, or those who appeal to a sense of 'general usage' without consulting a corpus of authentic texts.

The computer can be used not only to compute frequencies, but to view (using a concordancer) the contexts of a series of expressions. This can very efficiently establish the differing collocational patterns of words that are near synonyms, including the neologism / compound word pairs we saw above. For example, the difference between *disipi* (to waste) and *malŝpari* (to waste, 'to un-save') becomes clearer on the basis of contextual evidence based on the corpus. As can be seen above, both words have a roughly equal frequency in the corpus as a whole. But in terms of usage, *disipi* always in fact appears as a noun, collocating with words for strength or national resources (*disipo de fortoj* waste of strength*, disipo de riĉofontoj* waste of

sources of wealth, *disipo de monrimedoj* waste of wealth, *disipado de la fortostreĉoj* waste of resources) while *malŝpari* is used exclusively as a verb and is reserved for wasted time or breath (*malŝparis vortojn* wasted words, *malŝpariĝis la kritiko* criticism was wasted, *malŝparis siajn penojn* wasted their efforts, *malŝparas tempon* wastes time). Similarly, three similar adjectives *aperta* 'open' / *malfermita* 'open, unclosed' / *malferma* 'open, unclosed' are distinguished by a different range of collocations. *Aperta* is used figuratively in literary and technical contexts (*aperta kaj hidromekaniza elterigo* open-cast and hydraulic mining, *aperta faŭko* open jaws, *aperta monto-ĉeno* wide mountain range), *malfermita* is used for concrete objects with some aperture: *malfermita pordo* open door, *malfermita kameno* open chimney, *malfermita letero* open(ed) letter, while the simple form *malferma* (also 'open') essentially refers to public events: *malferma tago* open day, *malferma ceremonio* opening ceremony, *malferma letero* open letter (i.e. addressed to the public).

The distinction between single root synonyms or near-synonyms can also be demonstrated. For example, the Latinate *peni* (to try) is often perceived as close in meaning to the Slavic-derived *klopodi* (to strive). Although both can be distinguished semantically in terms of intensity, they also have distinct collocations which enable us to establish an overall pattern of usage or phraseology:

Peni	+ *serĉi novajn konceptojn*	To try	+ to search for new concepts
	+ *trovi la supron de pantalono*		+ to find the top of the trousers
	+ *misorienti la publikon*		+ to misdirect the public
	+ *ŝirmi per alta nomo mallaboremon*		+ to hide laziness high-mindedly
	+ *protekti nin kontraŭ fulmo*		+ to protect us from lightning
Klopodi	+ *helpi al tiu mondoparto*	To try	+ to help this part of the world
	+ *intertraktadi kun aliaj instancoj*		+ to interact with other bodies
	+ *respondi al ĉiuj*		+ to reply to all
	+ *vendi ĝin al la ŝtato*		+ to sell it to the state
	+ *inciti envion*		+ to incite envy

Even when the basic semantic difference is taken into account ('to try' and 'to strive' can only be approximate translations), the collocational patterns for each word provide a specific frame of reference for distinguishing the two words. *Peni* is consistently used to introduce processes of discovery (finding) or covering (to protect or hide), while *Klopodi* is consistently used with verbs denoting personal interaction

(selling, motivating, helping etc.). As with most collocational patterns in language, this is not an exclusive or 'rule-based' distinction, and the matter is somewhat more complicated than a mere difference in distribution of expressions. For instance, both verbs happen also to be used with the verbs *trovi* (to find), *montri* (to show) and *eldoni* (to publish). Insteda of weakening the phraseological patterns of these two verbs, these collocations can be interpreted as fitting into both general categories (revealing discovery as well as establishing contact and interaction). In addition, *peni* is also frequently preceded by adverbs serving to intensify its meaning (*insiste* insistently, *firme* firmly, *simple* simply, *vane* vainly) while *klopodi*, presumably because of its perceived strong semantic intensity is never accompanied by such expressions. The explanation for this difference in usage is unclear. However, this is at least consistent with findings on collocations in other languages, since collocational distinctions between related words often center on such features as negative / positive connotations, the presence or absence of emphatic modality, differing grammars of complementation and so on.

We have seen that the principle of phraseology extends to near-synonyms in the core vocabulary, but it can also distinguish between cognate forms across languages. In British English, the verbs 'persuade' and 'convince' have different grammatical collocations. 'Persuade' tends to be used with non-finite clauses expressing some action ('persuade them <u>to</u> go round the world') while 'convince' is used with complement finite clauses expressing some idea ('convince them <u>that</u> the world is round'). Again, it should be pointed out that this is not a rule or a distinction based on intuition: these patterns are tendencies observed from a corpus across thousands of examples (Sinclair 1991), and can only properly be established by a corpus analysis of millions of words. In French, we have also observed the distinction 'persuader / convaincre', and these takes a different phraseological direction ('persuader' tends to be used passively with 'that' clauses, while 'convaincre' tends to take direct objects and no further clauses). In Esperanto, our corpus shows that *persvadi* is like the English 'persuade' (...*persvadi nin forlasi la pruvitajn principojn* persuade us <u>to</u> discard proven principles, *ni povas persvadi eminentulojn veni* we can persuade eminent people <u>to</u> come, *li sukcesis ... persvadi la vendiston vendi al ili sufiĉe da glaciaĵo* he succeeded in persuading the shop keeper <u>to</u> sell them enough ice-cream). Conversely, *konvinki* is used in the same way as 'persuader' in French i.e. the passive form is

preponderant (*imperiismo, konvink<u>ita</u> pri sia nepuneblo* imperialism, convinced of its impunity, *ni estas konvink<u>itaj</u> pri ĝia kapablo...* we are convinced of its ability, *neniuj esperantistoj povas esti konvink<u>itaj</u> de tiu rezonado* no Esperantists can be convinced by that reasoning). The frequently used ergative *konvinkiĝi* is also comparable with passives in French and English (*mi konvinkiĝis ke la materialo estos...* I was convinced that the material would be..., *li per siaj propraj okuloj kaj oreloj konvinkiĝis* he was convinced by his own eyes and ears, *sed por tute konvinkiĝi, li komencis novan provon* but to really convince himself, he started a new experiment). This kind of evidence brings new light to the concept of near-synonyms, since clearly the communicative competence of the speaker is also expected to include some idea of the general tendency for words to be associated with global syntactic patterns. Esperantists are no different to other speakers, although what we have here is interesting evidence to suggest that phraseological patterns become fixed even in a short space of time (i.e. since 1887).

Collocational patterns also emerge in the distribution of morphemes and their derived forms. It is often suggested that different grammatical systems operate on a largely symmetrical basis, and different forms are often explained in terms of basic differences of meaning. This does not have to be the case, however. For example, as an independent word *iĝi* (to become) has a potentially similar role to the forms *estiĝi* (to come to be) and *fariĝi* (to be made). The verbs *estiĝi, fariĝi, iĝi* all appear to be the equivalent of 'become' in English, but many Esperantists feel that the different forms should correspond to a range of possible nuances:

> *ĝi iĝis blua* / *ĝi bluiĝis* 'it became blue'
> *ĝi estiĝis blua* 'it came to be blue'
> *ĝi fariĝis blua* 'it got blue'

The difference is very slight, and it may be a question of register or style rather than pure semantics. In any case, corpus analysis suggests that these forms have distinct distributions, and as the following concordances demonstrate, the difference between these forms is just as much a question of phraseology as it is a question of semantics (we suggest in fact that the choice *iĝi / estiĝi / fariĝi* is not as free as a semantic explanation would imply):

Iĝi ('to become, to get': 118 instances: Syntax: subject, verb, predicative attribute):

La nuntempa mondo <u>iĝis</u> tro malgranda	(today's world has become too small)
Pli riĉaj ankaŭ <u>iĝis</u> la sociaj fondusoj	(the social funds have become richer)
la unua afero <u>iĝis</u> publika temo…	(the first issue has become one of public debate)
<u>iĝas</u> pli drasta la sociala malegaleco	(social inequality becomes more drastic)
<u>Iĝas</u> malpli konvinkaj la provoj...	(the attempts […] get less convincing)

Estiĝi ('to develop, to come to be' 15 instances. Phraseologically: usually accompanied by explanations to problems. Syntactically: subject, verb, no predicative attribute):

Problemoj <u>estiĝas</u> pro tiu diferenco...	(problems develop because of that difference)
Pro malsana kruro <u>estiĝis</u> embolio	(because of a poorly leg, clotting developed)
Tiamaniere <u>estiĝis</u> la objektivaj kondiĉoj...	(thus objective conditions developed …)
la ŝoko kiun mi sentis ne <u>estiĝis</u> en ŝi...	(the shock I got did not affect (develop with) her)
certe ne <u>estiĝos</u> iu problemo alvertikaliĝi	(there will be no problem getting vertical)

Fariĝi ('to become': 76 instances. Syntax: as with a copular verb, the subject is replaced or transformed by the identifying / predicative complement):

Li <u>fariĝis</u> ebria de kolero	(he became drunk with anger)
ĝi <u>fariĝis</u> klara kiel kristalo	(it became clear as crystal)
Serĝento Kolimer <u>fariĝis</u> suboficiro	(Sergeant Kolimer became an N.C.O)
IKS <u>fariĝis</u> la zagreba kultura centro	(IKS became Zagreb's cultural center)
Tiu interpreto certe <u>fariĝos</u> leĝo	(that interpretation will certainly become law)

While the details of these patterns are interesting for each particular case, their consistency demonstrates the more general point that principles of phraseological patterning are at work even in such a seemingly free and arguably disparate language as Esperanto. Furthermore, even the most fluent speakers of English and French can not be expected to be consciously aware of the phraseologies of such specific patterns. The question remains as to where these patterns have come from, and how they became established across the wider community. While the distinctions between *persvadi* and *konvinki* or *peni* and *klopodi* may be down to the influence of foreign languages (and this is still not clear), the case of *iĝi* suggests that collocational patterns also emerge for expressions that are 'indigenous' to Esperanto and are associated with the established conventions of morphology in the language. None of

the donor languages exactly mirror the distribution of this morpheme, and there is no reason why any one ethnic language pattern should come to dominate a phraseology that is clearly so hidden from the conscious intuitions of its speakers. Such patterns can not be legislated or imposed, and we conjecture that they are more likely to emerge both as a function of the community of speakers and as functions of a language system that unknowingly attempts to consolidate global patterns of speech, much in the same way that the brain attempts to impose familiar structures onto vision.

In addition, collocations vary according to individuals and in different contexts and registers and this extends to patterns exhibited across time, different individuals and different genres. The language and style of individual authors has been a widely studied phenomenon in Esperanto. Kalocsay (1963:55-58) credits Grabowski's innovative style as having been one of the greatest influences on the language, in particular his translation of the Polish epic *Sinjoro Tadeo*. A number of Grabowskisms have become relatively established aspects of the language: in particular the free use of word class morphemes, the distinctive use of predicative verbs without the use of 'to be' (*prudenti* to be prudent, *preti* to be ready), compound participles (*malpermesitas* it is forbidden), complex verb compounds (*fruktuzi* to use fruitfully, *vortresumi* to resume in one word) and the unlimited use of simple and compound adverbs (*okule manĝi* to eat with the eyes, *minaci amasbuĉe* to threaten with mass butchery). Other authors have also pushed forward the boundaries of the language, either by liberally importing borrowed words or by throwing the language open to personal invention and fantasy as in the writings of Karolo Piĉo and the poetry of Valano. Other writers have capitalized on clarity of expression and originality without breaking the perceived boundaries of the existing language, including some of the best known exponents of Esperanto prose and poetry such as Auld (*La Infana Raso* The Child Race), Baghy (*Viktimoj* Victims), Matthias (*Fajron sentas mi interne* I feel a fire inside me) and.C. Rossetti (*Kredu min, Sinjorino* Believe me, my dear).

The study of genre and functional variation is less well appreciated than literary style although collocations and other phraseological patterns can be seen to change systematically from one context to the next, especially in scientific or journalistic

texts. Esperantists have been wary of attempts to identify national differences in among speakers (apart from accent), largely because characteristic speech patterns were possibly thought to undermine the universal ideal and were to be discouraged. The collocations we discussed above appear to demonstrate an opposite process however: subtle patterns of expression appear to have emerged despite national speech differences. Nevertheless, differences do occur, and we may attempt to establish some basic distinctions on the basis of the corpus evidence. Although our corpus is too small to distinguish between very general text types, the Russian *Soveta Esperantisto* (*S.E.*) comprises a distinctive segment for purposes of comparison. In Appendix 4, we have set out a general statistical comparison between *Soveta Esperantisto* and the corpus as a whole. Although lexical differences are immediately apparent, the distribution of grammatical items indicates more systematic differences of style. The list of words that are un-typical of the *S.E.* appears in Appendix 4.2. This shows that the general language (including fictional narratives and shorter more dialogic texts) is characterized by greater use of pronouns, conjunctions and a wider use of prepositions (apart from journalistic *en* and *pri* in the *S.E.*). Unsurprisingly, there is just more 'grammar' outside the *S.E.* than in (this is a statistical effect: only the most salient grammatical features of *S.E.* will be shown to be statistically significant at the top of the list). From 4.2, we can see that *Soveta Esperantisto* is characterized by a significantly higher use of the determiner *la* as well as *de*. Rather than indicating a national difference, both items are indicative of dense, highly modified and referential journalistic text. The word list comparison also confirms more specific stylistic tendencies of *S.E.*: a preference for the short ergative verb form *iĝi* (as opposed possibly to *estiĝi* and other forms ending in *iĝi*) as well as for the innovative agentive preposition *far*. Other patterns emerge lower down the list: *S.E.*'s preference for the use of the quality morpheme *–eco*: *kontraŭeco* (contrariness), *afereco* (affairs, 'general activity'), *ĝustatempeco* (just-in-time-ness), *organiziteco* (organized-ness). The *Soveta Esperantisto* also appears to use the morpheme *ej* in some unique ways: *entreprenejoj* firms, analogous to French 'entreprises', *konstruejoj* construction sites, *herbejoj* 'fields' where the term *kampo* is preferred in the rest of the corpus. *Soveta Esperantisto* is noticeably lacking in combined verbal participles (as in *malpermesitas* 'is forbidden') as well as other innovations, such as *rilate* + accusative. In the analysis of syntax above, we noted the use of grammatical constructions particular to Slavic, in particular the tendency to interrupt the determiner + noun structure of the noun phrase

with passives and prepositional phrases. This principle extends to other longer modifying structures that can be found relatively frequently in *S.E.*: *oni eldonas per neimageblaj por la mondo eldonkvantoj la klasikan kaj modernan poezion …* 'one publishes [in unimaginable for the world quantities] classical and modern poetry' (in correct English word order: 'one publishes classical and modern poetry in quantities (that are) for the world unimaginable'). This formulation occurs elsewhere in the corpus, for example in Zamenhof's letters, the active participle is used in a non-finite clause which also intervenes between the determiner and adjective: *la frapinta ilin neordinara facileco de la lingvo*: 'the striking them extraordinary easiness of the language' (this is however a translation from Russian).

This kind of comparison is too specific to *Soveta Esperantisto* to encompass systematically the differences between other textual genres in the language, not least those of narrative structure and rhetorical function. Nevertheless, the corpus-based methodology demonstrates that differences can be identified in unexpected areas and that variation extends across all the traditional levels of language use (syntax and morphology as well as phraseology). In time, it should be possible to fully characterize the phraseological features of several different registers in Esperanto (for example, the writings of Zamenhof or other writers compared with general Esperanto, or the writing of pre-war Esperantists compared with modern writers).

4.3 Cultural reference

Since Esperantism is far from being a traditional 'speech community', it may be more appropriate to use the term 'discourse community' proposed by Swales (1990). The discourse community is a group defined by its own use of language, most often specialists (Swales was originally describing scientists), who have a clearly defined purpose, and who develop their own social hierarchies, networks of communication and their own technical jargon. Membership of a speech community is usually involuntary and difficult to obtain, while membership of a discourse community is voluntary and temporary, and involves membership of other communities as well. While speech communities are defined on geographic or dialectal grounds, the discourse community is defined by topic or discipline and by the extent to which

members are familiar with the main protagonists, arguments and issues relevant to that discourse (a similar position is taken by Fettes 1996). It is possible to be a poor speaker of a language, but fluent in its discourse of politics or commerce (many non English-speaking scientists use English in scientific discourse but this does not on its own make them members of a speech community, and members of the speech community of English are not fluent in the discourse of science). Esperantists can thus be partly defined by the traditional speech-community criterion of fluency (some individuals do have the status of native speakers), but they can also be usefully defined in terms of their knowledge of the discourse of Esperantism. Thus we can distinguish between amateurs, interlopers and the 'insiders' who are at the heart of debate in the Esperanto movement, who know the key issues and individuals who have a place in or are recognized as members of the movement. Much debate in Esperantism is currently directed at the definition of the movement's purpose, and while there are counter-currents (for example idealists versus pragmatists, reformists versus conservatives) they are at least familiar with each other's modes of speech and argument: a prerequisite for membership of the discourse community.

The jargon of a discourse community is clearly the most tangible kind of phraseology, and any fluent Esperantist is defined by his or her knowledge of the terms that have been developed. The Esperanto culture is preoccupied by the idea of community, as can be seen in the large number of words relating to group membership: *denaskulo-* a mother-tongue speaker, *esperantujo–* the community of Esperantists, *eterna komencanto-* a perpetual learner, *varbito-* recruit, *samideano–* ideological comrade*, movadano, movadulo–* an insider or activist, *kabeisto–* someone who defects or disappears (after *Kabe,* the nickname of a famous defector), *fina venko-* the final victory (a reference to the adoption of Esperanto), *fina venkisto-* 'hardline Esperantist', *krokodilo–* someone who rudely speaks another language at meetings, *aligatoro* someone who rudely speaks someone else's language at meetings, *idisto* – a schismatic, *volapukisto-* speaker of Volapük (speakers of rival projects are considered legitimate victims for Esperantist jokes) and so on. There are movement-based terms for currents, schools of thought, trends, ideologies and methods. Here is a short selection:

analiza skolo - an attempt to establish and apply the original principles of the language strictly, including transliterating foreign words (propagated by R Schulz).

atismo - supporter of the aspectual view of past participles.

Cseh skolo - one of a number of internationally-based language-teaching methods.

fina-venkismo – refers to 'the final victory', the ideal of having Esperanto adopted as an official second language in place of any national language.

Hamburga pûĉo - Hamburg Putsch, the scandalous replacement of the UEA President Ivo Lapenna in 1974.

idista skismo – the split in the movement provoked by the publication of the reform project *Ido* in 1907.

interna ideo - the 'internal idea', the anti-nationalist ideology underpinning Zamenhof's motivation for inventing the language.

One term, *raŭmismo,* represents one of the more recent controversies in the Esperantist movement. The term refers to a series of policy statements published in the Finnish town of Raumo by J. Lindstedt and writers involved in the literary review *Literatura Foiro* in the 1980s. *Raŭmismo* represented an attempt to steer Esperantism away from a preoccupation with '*fina-venkismo*' towards a more cultural community based around a neutral alternative world-culture. The move was criticized by those Esperantists who emphasized Esperanto's role in the defence of minority languages and cultures (Piron 1994) but has become a easily recognizable term to refer to the various political tendencies within the movement.

In addition to movement-based terms, a number of artefacts and cultural symbols are also associated with the movement and require some explanation to outsiders: *La Espero*: the Esperantist anthem with its own official musical score, *la verda stelo* – the green star of hope, adopted as the Esperantist symbol used on badges and flags, *speso* – Esperanto money (*spesoj* were actually minted as an international currency and the term is still referred to in stories). Recent editions of *Asterix* and *Tintin* produced in Esperanto include several cultural references to Esperantism and in the usual parade of in-jokes the translators relentlessly parody Zamenhof's proverbs: *ne ŝovu la nazon en fremdan amforon* 'don't put your nose in a foreign amphora', *de mano al buŝo disverŝiĝas la supo* 'from hand to mouth the soup gets spilt' or refer to Esperanto hymns and poetry (the galley slaves sing songs by the writer Valano, and the Bard tries to sing *En la mondon...*the first phrase of *La Espero*).Other preoccupations have led to simple expressions being used more or less idiomatically

as special movement-related terms: *eldoni* 'to publish', give out publications, *aliĝi* 'to join', to sign up to a conference, *estrarano* 'board member', member of the main U.E.A committee. Finally, the large number of translations of important but stylistically divergent works (from Shakespeare's sonnets to Tom Sawyer) have all created the need for complex systems of register in Esperanto, and these artifices have naturally begun to impact on the general language. Borrowings from translation and neologisms in Esperanto's original works constitute an important resource for terminology, and are designed to have some colloquial, slang or archaic meaning, some examples being *drugo* ('mate, pal' from Russian), *kasao* ('house, digs' from Spanish), *kaputa* ('bust, knackered' from German) (Halvelik 1973). These terms are being deliberately chosen and changed to fill a perceived stylistic gap, a slightly different function to words taken on for new technical terms. The fact that these words are chosen and become generally recognized (if not actively used) is key evidence of the existence of a community and also of dynamic change in the language.

Attitudes to language itself form an integral part of the Esperanto culture. Perhaps in reaction to the 19th century preoccupation with the genius of the national language, many Esperantists played down the idea of idioms and condemned the irregularities of the national languages, emphasizing instead Esperanto's ability to absorb and integrate a wide selection of national literatures. It can be argued however that idioms, formulaic expressions, stylistic variants and other phenomena obey consistent principles of their own and are an inevitable part of a language system. However, this is not a widely held view, and it is common for commentators to attempt to censure each other's expressions. A minority of Esperantists claim that they can preserve and even improve the logical structure of the language. For example, proponents of the *analiza skolo* as well as individuals such as Bernard Golden have regularly attacked Kalocsay and Waringhien's descriptive approach for allegedly creating illogical forms and for disrupting the perceived integrity of the language. They were especially charged with exposing Esperanto to subversive literary innovation and were attacked for attempting to justify 'foreign' forms. Descriptive attitudes to language are often criticized in the letters pages of the Esperanto press, and this is the same as prescriptive conservatism in other language communities. To be fair, much prescriptive work is usefully aimed at learners' mistakes (Venture 1977), and the conservatives' arguments are often based on firm, logical analogy. However, there is

no evidence to suggest that prescription or censure have any effect on the language's general development and, of course, many Esperantists have argued a different case, for a more descriptive and tolerant perspective (Auld 1986, Duc Goninaz 1988, Janton 1994).

The idea of 'irregularity' in Esperanto needs especially to be reassessed, as Duc-Goninaz (1988) has argued. The popular and conservative conception of Esperanto has often been premised on the idea that language constitutes a single homogenous system with a hidden logic underpinning it. But this is in direct contradiction to the work of many linguists (especially followers of Firth, but also some generativists and neurolinguists) who are proposing a different notion of language as essentially a collection of independent modules, often competing and providing for useful redundancy in the system. The descriptive approach we have taken in this guide has attempted to demonstrate some of the general tensions that these systems may create in Esperanto, and to provide some evidence to suggest that they are an inevitable aspect of human communication and worthy of more attention.

5 INTERLINEAR TRANSCRIPTION.

1)   *Vid-handikap-it-o-j*   *pov-as*
see-handicap-PASSIVE PART-N-PL   can-PRES
abon-i   *la*   *revu-o-n*   *Esperanto*   *sur*
subscribe-INF   the   review-N-ACC   Esperanto   on
son-kased-o.
audio-cassette-N.

'The visually handicapped can subscribe to the review 'Esperanto' on audio tape'.

2)   *Kun-e*   *kun*   *la*   *verdikt-o*   *aper-as*
With-ADV   with   the   verdict-N   appear-PRES
la   *nom-o-j*   *de*   *la*   *juĝ-int-o-j.*
the   name-N-PL   of   the   judge-PERF PART-N-PL.

'Together with the verdict appear the names of judges'.

3)   *La*   *nun-a-j*   *indik-o-j*   *atest-as*   *ke*
The   now-ADJ-PL indication-N-PL   attest-PRES   that (conj.)
trov-iĝ-as   *Esperanto-bibliotek-o-j*   *en*   *tut-a*
find-ERG-PRES   Esperanto-library-N-PL   in   whole-ADJ
Eŭrop-o
Europe-N.

'Current indications show that Esperanto-libraries are found throughout Europe'.

4)   *Ni*   *asert-as*   *ke*   *lingv-a*   *mal-egal-ec-o*
We   assert-PRES   that (conj.)   language-ADJ un-equal-ABS-N
sekv-ig-as   *komunik-a-n*   *mal-egal-ec-o-n*
follow-CAU-PRES.   communicate-ADJ-N   un-equal-ABS-N-ACC
je   *ĉi-u-j*   *nivel-o-j,*   *inkluziv-e*   *de*
[ ]-PREP   all (COR-DET-PL)   level-N-PL   inclusive-ADV   of
la   *inter-naci-a*   *nivel-o*
the   inter-nation-ADJ   level-N.

'We assert that linguistic inequality leads to communicative inequality at all levels, including the international level'

ABBREVIATIONS:

General:

UEA = *Universala Esperanto Asocio.*   S.E. = *Soveta Esperantisto.*

Functional Syntax:

Sb	= Subject	Complements:
Vb	= Verb	Ci = Indirect / Prepositional
Ob	= Complement (Direct Object)	Ca = Predicative / Attributive
Ab	= Adverb, Adjunct	Ce = Equitive / Identifying
Cj.	= Conjunction	// = Modifier / clause boundary.
Dt	= Determiner	[ ] = Embedded phrase or clause.

Structural Syntax:

IP = Sentence / Inflection Phrase
NP = Noun Phrase
VP = Verb Phrase
PP = Prepositional Phrase
AP = Adjective Phrase
AdvP = Adverb Phrase
D = Determiner

Morphology:

-	= Morpheme boundary	INF. = Infinitive verb marker
ABS	= Abstract quality morpheme	N = Nominal morpheme
ACC	=Accusative case	PART. = Participle
ADJ	= Adjective morpheme	PAST = Past tense marker
ADV	= Adverb morpheme	PERF = Perfective
CAU	= Causative voice morpheme	PL = Plural
COR	= Correlative prefix	PRES = Present tense marker
DET	= Correlative suffix 'determiner'	PROG = Progressive
		[ ]-PREP = Empty preposition
ERG	= Ergative morpheme	

REFERENCES

Auld, William. 1976. 'The Development of Poetic Language in Esperanto.' *Esperanto Documents* No.4A, Rotterdam and London, Centro de Esploro kaj Dokumentado pri la Monda Lingvo-Problemo (CED).

Auld, William. 1986. *Kulturo kaj internacia vivo*. Chapecó-Brazil, Fonto.

Bañeres, Jordi and Strubell, Miquel. 1998. *Discussion Manual on Lesser-Used Languages*. Brussels, EBMUL.

Bastien, Louis. 1909*1950. *Naŭlingva Etimologia Leksikono*. London, Esperanto Publishing Company.

Blanke, Detlev. 1985. *Internationale Planschpachen*. Berlin, Akademie-Verlag.

Butler, Montagu C. 1932. *Proverbs in Esperanto and English.* London, Leopold Hill.

Cherpillod, André. 1989. *Les langues agglutinantes et l'Espéranto*. Courgenard - France, Author's edition.

Cherpillod, André. 1988. *Nepivaj vortoj*. Courgenard - France, Author's edition.

Corsetti, Renato. 1987. *Knedu min sinjorino!* Pierrefitte-France, Kancerkliniko.

Cresswell J. and J. Hartley (revised by J. H. Sullivan). 1992. *Teach Yourself Esperanto*. Chicago, NTC Publishing Group.

Cruse, David. 1984. *Lexical Semantics*. Cambridge, Cambridge University Press.

Dashgupta, Probal. 1993. 'Idiomaticity and Esperanto texts: an empirical study.' In *Linguistics* 31/2: 367-386.

Drezen, Ernest. 1930*1972. *Analiza historio de Esperanto-movado'*. Leipzig, Eldon-Kooperativo por Revolucia Esperanto-Literaturo.

Duc-Goninaz, Michel. 1988. 'L'hétéroclisie- La règle et l'exception en Espéranto.' In *Travaux du Cercle Linguistique*. 6. Aix-en-Provence, Université d'Aix-Marseille III.

Duličenko, Aleksandr Dmitrievič. 1990. *Meždunarodnye vspomogatel'nye jazyki.* Tallinn: Valgus.

Eco, Umberto. 1997. *The Search for the Perfect Language.* London, Fontana.

Edwards, John. 1994. *Multilingualism.* London, Penguin.

Edwards, John. & MacPherson Lynne. 1987. 'Views of Constructed Languages, with Special Reference to Esperanto: An Experimental Study' in *Language Problems and Language Planning* 11: 283-304.

Fernando, Chintra 1996. *Idioms and Idiomaticity.* Oxford, Oxford University Press.

Fettes, Mark. 1996. 'The Esperanto Community: A Quasi-Ethnic Linguistic Minority?' In *Language Problems and Language Planning.* 20/1: 53-59.

Fielder, Sabine. 1999. 'Phraseology in Planned Languages: an Empirical Survey'. In *Language Problems and Language Planning.* 23/2:

Forster, Peter. 1982. *The Esperanto Movement.* Den Haag, Mouton.

Gledhill, Christopher. 1999. 'Towards a phraseology of English and French'. In C. Beedham (ed.) *Language and Parole in Synchronic and Diachronic Perspective.* Societas Linguistica Europaea. Oxford: Pergamon: 221-37.

Gopsill, F.P. 1988. *Short Interlingua Grammar and Vocabulary.* Bebbington-England, Author's edition.

Green, John. 1988. 'Romance creoles.' In Martin Harris & Nigel Vincent (eds.) *The Romance Languages.* London, Croom Helm: 420-473.

Grosjean-Maupin E., Esselin A., Grenkamp-Kornfeld S. & G. Waringhien. 1934*1988. *Plena Vortaro de Esperanto.* (10th edition). Paris, Sennacieca Asocio Tutmonda.

Hagège, Claude. 1985. *L'homme de paroles.* Paris, Fayard.

Halliday, Michael A. K. 1994. *Introduction to Functional Grammar.* London, Edward Arnold.

Halvelik, Manuel. 1973. *Popilo: popola idiomo, la verda dialekto.* Bruges, Sonorilo.

Haupenthal R. (ed.) 1978 *Eugen Wüster. Esperantologiaj Studoj: Memora Kolekto.* Antwerp and Laguna, Stafeto and T. Kehler.

Janton, Pierre. 1994. *L'Espéranto. Que sais-je?*. (4th Edition). Paris, Presses Universitaires de France.

Jordan, David K. 1997. 'Esperanto and Esperantism: Symbols and Motivations in a Movement for Linguistic Equality'. In Tonkin (ed.) 1997: 39-65.

Kalocsay, Kálmán: 1931*1963. *Lingvo, Stilo, Formo*. Budapest, Librejo Pirato.

Kalocsay, Kálmán & Gaston Waringhien. 1985. (5th Edition). *Plena analiza gramatiko de Esperanto*. Rotterdam, Universala Esperanto Asocio.

Lapenna, Ivo, Lins, Ulrich & Carlevaro Tazio. 1974. *Esperanto en perspektivo*. London, Centro de Esploro kaj Dokumentado pri la Monda Lingvo-Problemo (CED).

Large, Andrew. 1985. *The Artificial Language Movement*. Oxford, Basil Blackwell.

Lins, Ulrich. 1990. *La danĝera lingvo*. Moscow, Progreso. (trans. 1988. *Die Gefährliche Sprache*. Gerlingen, Bleicher Verlag.)

Lo Jacomo, François. 1982. *Liberté ou autorité dans l'évolution de l'espéranto?* Pisa, Edistudio.

Mattos, Geraldo. 1987. *La deveno de Esperanto*. Chapéco-Brazil, Fonto.

Migliorini, Bruno. 1985. *Lingvaj aspektoj de Esperanto (1924-1963). Eseoj*. Edistudio, Pisa.

Papegaaij W. and Schubert R. 1988 *A Corpus-based Bilingual Knowledge Bank for Distributed Language Translation* Amsterdam, DLT Publications.

Pei, Mario. 1973 Artificial languages: international (auxiliary). In *Current Trends in Linguistics XII*. :999-1017. The Hague, Mouton.

Pirlot, G. 1982. 'The Official Situation of Esperanto Instruction in the World' in R. Eichholz and V. Eichholz (eds.) *Esperanto in the Modern World.* Bailieboro- Canada, Esperanto Press.

Piron, Claude. 1989a. 'A few notes on the evolution of Esperanto' in Schubert (ed.) 1989: 129-143.

Piron, Claude 1989b 'Who are the speakers of Esperanto?' in Schubert (ed.) 1989: 157-172.

Piron, Claude. 1991. 'Esperanto: European or Asiatic Language?' in *Esperanto Documents*. 22/A. Rotterdam, Universala Esperanto Asocio.

Piron, Claude. 1994. *Le défi des langues: du gâchis au bon sens*. Paris, L'Harmattan.

Privat, Edmond. (trans. Ralph Eliott). 1920*1931. *The Life of Zamenhof*. London, George, Allen & Unwin.

Rasič, Nikola (ed.). 1994. *La rondo familia. Sociologiaj studoj en Esperantio*. Pisa, Edistudio.

Rasmussen, Anne. 1997. 'Langue internationale: histoire d'une quête (1880-1914)' in *Pour la Science*. October: 116-118.

Saussure, René de. 1910. *La construction logique des mots en espéranto*. Geneva, Universala Esperanto-Librejo.

Schubert, Klaus. 1993 'Semantic Compositionality. Esperanto word formation for language technology'. In *Linguistics* 31/2: 311-365.

Schubert, Klaus (ed). 1989a. *Interlinguistics. Aspects of the Science of Planned Languages*. Berlin: Mouton de Gruyter.

Schubert, Klaus 1989b. 'An unplanned development in planned languages.' in Schubert (ed.) 1989: 249-275.

Sherwood, Bruce Arne 1982. 'Statistical analysis of conversational Esperanto, with discussion of the accusative.' *Studies in the Linguistic Sciences*. 12: 165-182.

Sinclair, John McH. 1991. *Corpus, Concordance, Collocation*. Oxford, Oxford University Press.

Stocker, Frank. 1996. *Wer spricht Esperanto? Kiu parolas Esperanton?* München, Lincom Europa.

Swales, John. 1990. *Genre Analysis*. Cambridge, Cambridge University Press.

Tonkin, Humphry (ed.). 1997a. *Esperanto, Interlinguistics and Planned Languages*. New York, University Press of America.

Tonkin, Humphrey. 1997b. 'One Hundred Years of Esperanto: A Survey'. In Tonkin (ed.) 1997a: 71-92.

Venture, Alec. 1977. *Laŭokazaj lingvaj notoj.* London, Brita Esperanto Asocio.

Versteegh, Kees. 1991 'The substratum debate in creole linguistics.' In *Diachronica.* 8: 59-80.

Versteegh, Kees 1993 'Esperanto as a first language: Language acquisition with a restricted input.' In *Linguistics.* 31/3: 539-555.

Veuthey, Francisco L. 1996 *UEA Jarlibro.* Rotterdam, Universala Esperanto Asocio.

Vilborg, Ebbe 1995 *Etimologia Vortaro de Esperanto* Vol. 1-4. Malmö, Eldona Societo Esperanto.

Wacha, Balázs. (1986). 'Esperanto kaj ĝia uzado en familio. In *Kontribuo al Lingvaj Teorio kaj Praktiko.* IV: 61-73. Poprad, Slovaka Esperanto-Asocio.

Wackrill, Alfred E. 1907 *Konkordanco de Ekzercaro.* Paris, Hachette.

Waringhien, Gaston. 1970 (1981, 1987) *Plena Ilustrita Vortaro.* Paris, Sennacieca Asocio Tutmonda.

Wells, John C. 1969. *Concise Esperanto and English Dictionary.* London, Hodder and Stoughton.

Wells, John C. 1989. (2nd Edition). *Lingvaj aspektoj de Esperanto.* Rotterdam, Universala Esperanto Asocio.

Woolf, Felix. 1982. *English Phrases and Expressions in Esperanto.* Balieboro-Canada, Esperanto Press.

Zamenhof, Ludwik L. 1887. *Internacia lingvo: Antaŭparolo kaj plena lernolibro por rusoj.* Warsaw, Kelter Lithographs. (reprinted in Huĝimoto et al. (eds.) *Unuaj libroj por esperantistoj.* Kyoto, Eldonejo Ludovikito, Vol. 1.1973).

Zamenhof, Ludwik L. 1962. 'Lingvaj respondoj' published in Gaston Waringhien (ed.): *Dr. L. L. Zamenhof. Lingvaj respondoj. Konsiloj kaj opinioj pri Esperanto.* Marmande, Editions françaises d'Espéranto.

APPENDICES: CORPUS DATA.

Corpus size: 1 563 500 running words (first edition 312 130)
Types (different word forms): 37 435.   Texts used: 156.
Mean word length: 4.9 letters.   Mean sentence length: 11.42 words.

All texts were analyzed using an early version of the Wordsmith linguistic statistics package (available commercially on the Internet from: http://www.hit.uib.no/wordsmith/).

Composition: The corpus is very small by current standards but represents a broad selection of texts. It is effectively a 'sample corpus'. Larger text corpora are needed to study specific lexical items and to compare genres (journalism vs. narrative, spoken political discourse vs. written political discourse etc.). A small corpus of Esperanto can only hope to establish global grammatical patterns. Other corpus work on Esperanto includes the Distributed Language Translation Project (reported in Tonkin 1997a) and spoken language projects reported in Schubert (1989a).

The corpus includes a wide range of registers: journalistic texts (reviews and reports from the magazine *Esperanto* and the national Esperanto magazines, political tracts from *Soveta Esperantisto*, conference reports, articles on current affairs and on language etc.), legalistic and administrative texts (the Prague Manifesto, statements from Amnesty International, instructions and regulations from various organizations), literary works (including poems by Zamenhof and others, novels, short stories) and personal internet sites in Esperanto (informal letters, personal statements, propaganda). The vocabulary list demonstrates the influence of the pieces coming from 'Soveta Esperantisto' (some 80 000 words in length). Novels and short stories account for over half the corpus: 870 000 words. A number of translated texts were included, although this is not normal practice (for example, the Oxford-based British National Corpus excludes the Bible and even widely available translations, such as Asterix on the grounds that these texts may affect the authenticity of the English contained in the corpus). However, given the non-native nature of Esperanto and the normative status of some of the texts included in our corpus (e.g. short extracts from the Bible, Zamenhof's translations as well as translations of Zamenhof's correspondence in various languages), the inclusion of these texts can in some way be justified until a wider set of texts become available.

Appendix 1: Word Frequency Lists: 1 First 100 Items.

RANK / WORD	/ Frequency /	(% of the corpus)	RANK / WORD	/ Frequency /	(% of the corpus)
1 LA (the)	25528	(8.2%)	51 ESPERANTO	542	(0.2%)
2 DE (of, from)	12949	(4.1%)	52 TION (that, *acc.*)	541	(0.2%)
3 KAJ (and)	10046	(3.2%)	53 DO (then, thus)	538	(0.2%)
4 EN (in)	5685	(1.8%)	54 CHIUJ (all those)	537	(0.2%)
5 AL (to, towards)	4472	(1.4%)	55 TIUJ (those)	533	(0.2%)
6 MI (me, I)	4235	(1.4%)	56 MIN (me, *acc.*)	531	(0.2%)
7 ESTAS (am, is, are)	3804	(1.2%)	57 GHIN (it, *acc.*)	518	(0.2%)
8 NE (no, not)	3713	(1.2%)	58 HAVAS (has, have)	507	(0.2%)
9 POR (for)	2925	(0.9%)	59 JE (*empty prep.*)	503	(0.2%)
10 LI (he)	2701	(0.9%)	60 TIEL (so)	500	(0.2%)
11 KE (*conj.* that)	2633	(0.8%)	61 POST (after)	483	(0.2%)
12 PRI (about)	2436	(0.8%)	62 OL (than)	481	(0.2%)
13 VI (you)	1918	(0.6%)	63 ANTAU (before)	473	(0.2%)
14 NI (we)	1910	(0.6%)	64 TIE (there)	457	(0.1%)
15 SED (but)	1908	(0.6%)	65 DEVAS (must)	456	(0.1%)
16 ESTIS (was, were)	1675	(0.5%)	66 SIA (*reflex det.*)	435	(0.1%)
17 ILI (they)	1500	(0.5%)	67 INTER (between)	422	(0.1%)
18 KUN (with)	1394	(0.4%)	68 ALIAJ (others)	420	(0.1%)
19 KIU (who)	1390	(0.4%)	69 LIN (him, *acc.*)	413	(0.1%)
20 TIU (*det.* that)	1330	(0.4%)	70 LAU (according to)	394	(0.1%)
21 TIO (*pron.* that)	1229	(0.4%)	71 TAMEN (however)	390	(0.1%)
22 EL (out of)	1207	(0.4%)	72 ECH (even)	387	(0.1%)
23 PLI (more)	1201	(0.4%)	73 SIN (*reflex pron. acc.*)	383	(0.1%)
24 KIEL (as)	1193	(0.4%)	74 DU (two)	379	(0.1%)
25 ONI (*pron.* one)	1178	(0.4%)	75 SIAN (*reflex acc.*)	369	(0.1%)
26 ANKAU (also)	1026	(0.3%)	76 MEM (*emphatic*)	360	(0.1%)
27 GHI (it)	986	(0.3%)	77 TRE (very)	359	(0.1%)
28 PER (by, with)	936	(0.3%)	78 LIA (his)	349	(0.1%)
29 SUR (on)	856	(0.3%)	79 TUTE (completely)	348	(0.1%)
30 SE (if)	843	(0.3%)	80 GHIS (until)	345	(0.1%)
31 CHI (*dem.* this)	834	(0.3%)	81 TIUN (that *acc.*)	342	(0.1%)
32 NUR (only)	786	(0.3%)	82 KION (what *acc.*)	339	(0.1%)
33 DUM (during)	767	(0.2%)	83 ESTI (to be)	338	(0.1%)
34 DIRIS (said)	760	(0.2%)	84 KIO (what)	338	(0.1%)
35 POVAS (can)	743	(0.2%)	85 HOMOJ (people)	337	(0.1%)
36 DA (*quant.* of)	741	(0.2%)	86 KIUN (which, *acc.*)	336	(0.1%)
37 SHI (she)	687	(0.2%)	87 MONDO (world)	332	(0.1%)
38 KIUJ (which, *pl.*)	667	(0.2%)	88 ANKORAU (*adv.* still)	329	(0.1%)
39 CHU (*interrog.*)	659	(0.2%)	89 KONTRAU (against)	323	(0.1%)
40 UNU (*num.* one)	651	(0.2%)	90 CHIU (every)	315	(0.1%)
41 KIAM (when)	643	(0.2%)	91 TIAM (then)	311	
42 NUN (now)	641	(0.2%)	92 JA (indeed)	305	
43 AU (or)	634	(0.2%)	93 GRANDA (big, large)	302	
44 PRO (because of)	622	(0.2%)	94 EBLE (possibly)	301	
45 CHAR (because)	621	(0.2%)	95 KIE (where)	296	
46 JAM (already)	602	(0.2%)	96 JAROJ (years)	287	
47 NIA (our)	599	(0.2%)	97 JEN (here is…)	287	
48 MIA (my)	579	(0.2%)	98 IU (some one)	284	
49 CHE (at)	571	(0.2%)	99 LANDOJ (countries)	283	
50 PLEJ (most)	555	(0.2%)	100 SCIAS (knows)	282	

Appendix 1.2: Word Frequency List (First 100 Lexical Items).

RANK / WORD	*/ Frequency*	*RANK / WORD*	*/ Frequency*
1 HAVAS (have)	507	52 PATRO (father)	164
2 DEVAS (must)	456	53 OKAZIS (happened)	160
3 ALIAJ (others)	420	54 RESPONDIS (answered)	160
4 ESPERANTO	542	55 EKZISTAS (exists)	139
5 TUTE (completely)	348	56 DENOVE (again)	159
6 ESTI (to be)	338	57 SOVETIA (Soviet)	156
7 HOMOJ (people)	337	58 GHUSTA (exact)	152
8 MONDO (world)	332	59 USSR	152
9 GRANDA (big)	303	60 ESTU (be, *volative*)	150
10 LANDOJ (countries)	283	61 POLITIKO (policy)	147
11 SCIAS (know, knows)	282	62 AFERO (matter)	146
12 ALIA (other)	273	63 SAME (likewise)	146
13 MULTAJ (many)	267	64 PROBLEMOJ (problems)	141
14 LANDO (country)	266	65 EUROPO (Europe)	137
15 TUTA (complete)	266	66 JUNA (young)	136
16 ESTOS (will be)	252	67 NU (*interj.* well)	138
17 BONE (well)	250	68 POLITIKA (political)	135
18 PARTIO (party)	245	69 FARAS (do, does)	131
19 LABORO (work)	238	70 DIVERSAJ (several)	130
20 KELKAJ (some)	235	71 EBLAS (is possible)	130
21 LINGVO (language)	230	72 DIRI (to say)	129
22 UNUA (first)	229	73 EVOLUO (evolution)	128
23 USONO (U.S.America)	229	74 KONGRESO (congress)	127
24 INTERNACIA	227	75 VENAS (come, comes)	126
25 DIO (God)	225	76 BONA (good)	125
26 MULTE (much)	225	77 JARO (year)	124
27 NOVA (new)	225	78 VIRO (man)	124
28 TEMPO (time)	224	79 EKONOMIA (economic)	123
29 POVIS (could be)	220	80 SOCIA (social)	123
30 VOLAS (want, wants)	218	81 SOCIO (society)	123
31 VIVO (life)	214	82 VIDAS (see, sees)	123
32 FARI (to do)	213	83 DIA (Godly)	121
33 ESTUS (would be)	210	84 SHTATOJ (states)	120
34 HOMO (a person)	207	85 PAROLIS (spoke)	118
35 SOVIETA (soviet)	207	86 DEVIS (had to, must)	117
36 VENIS (came)	203	87 MEMBROJ (members)	117
37 HAVIS (had)	197	88 ARMILARO (armaments)	116
38 DIRIS (said)	192	89 JUNULARA (youth movement)	116
39 VERE (really)	190	90 SOCIALISMA (Socialist)	115
40 VIDIS (saw)	190	91 TEMAS (is about)	115
41 MILITO (war)	189	92 ATINGI (to reach)	114
42 UNIO (union)	187	93 EKONOMIO (economy)	114
43 JUNULARO (youth movement)	178	94 RILATOJ (relations)	114
44 NECESAS (needs)	178	95 DONIS (gave)	113
45 POPOLO (the people)	178	96 KOMENCIS (began)	113
46 CERTE (certainly)	176	97 TUTAN (all of, whole, *acc.*)	111
47 MOVADO (movement)	173	98 KREDAS (believes)	111
48 FARIS (did, made)	171	99 EKZEMPLE (e.g.)	110
49 OFTE (often)	168	100 SINJORO (Mr.)	110
50 PACO (peace)	168		
51 POVUS (might be)	165		

Appendix 2. Sample-based phoneme survey*

a	20 108	b	4 732	c	8 265	ĉ	5 945
d	24 417	e	23 889	f	7 976	g	6 951
ĝ	2 843	h	4 517	ĥ	28	i	9 507
j	4 957	ĵ	307	k	35 101	l	36 815
m	17 852	n	14 344	o	5 075	p	26 311
r	8 287	s	23 164	ŝ	2 523	t	15 295
u	4 169	ŭ	120	v	1 452	z	373

*Sample size: 100 000 words.

Appendix 3 Distribution of word classes*

	Singul.	Plural	Sing. Accus.	Plural Accus.			Total	% of Corpus
Nouns	41261	14811	11606	4314			**71992**	23.3%
Adjectives	23525	9397	5510	2503			**40935**	13.3%
	Normal	Accus.						
Pronouns	19281	3934					**23215**	7.5%
Adverbs	15382	672					**16054**	5.2%
	Infinit.	Present	Past	Volitiv.	Future	Condit.		
Verbs	10362	15471	14838	12300	2226	1163	**56360**	18.3%
Functional adverbs	12951						**12951**	4.2%
Determiners	30184						**30184**	9.8%
Prepositions	38781						**38781**	12.6%
Conjunctions	17821						**17821**	5.8%
							308293	*100%*

*(Figures from the smaller first edition corpus. Excluding numerals, proper nouns and unclassified particles. This category includes 3837 items, approximately 1% of the overall corpus).

Appendix 4 . Word list comparison: *Soveta Esperantisto* vs. *General Esperanto*.

The following statistics are all from the first edition. They set out the general lexical characteristics of the *Soveta Esperantisto* magazine as they compare with our general corpus as a whole. The 'wordlist' computer program compiles and then compares two word frequency lists for each corpus and then calculates the Chi square (X2) significance of the words in the *Soveta Esperantisto*. The program then places the most statistically significant words in *Soveta Esperantisto* at the top of the list (a score of p = 0.000 or less is very highly significant). The resultant list allows the linguist to identify not only the most typical lexical items but also the most significantly salient grammatical items associated with the style of the *Soveta Esperantisto* journal. The program is part of the general 'Wordsmith' suite, as detailed above, available commercially from Oxford University Press [NB. Some non-lexical symbols are misread by the program and this accounts for the presence of any 'typing' errors in the lists].

Appendix 4.1 Typical Items in the *Soveta Esperantisto*:

Here we present the first 100 items, and also the most significant grammatical items within the first 1000 significant items. Significance (or 'typicality') decreases down the list. For example, the word *politiko* (politics, policy) occurs so frequently in the *Soveta Esperantisto* (127 times compared with a corpus total of just 147) that it is calculated as being the 15th most significantly typical word in the S.E. subcorpus.

	SOVETA ESPERANTISTO			ESPERANTO CORPUS			
	WORD	Freq.	%	Freq.	%	X2	p
1	DE	6405	(7.8%)	12949	(4.1%)	1819.1	0.000
2	LA	8485	(10.3%)	25528	(8.2%)	362.7	0.000
3	NUKLEA	209	(0.3%)	210		210.7	0.000
4	SOVETA	207	(0.3%)	207		209.9	0.000
5	UNIO	174	(0.2%)	187		160.7	0.000
6	SOVETIA	156	(0.2%)	156		157.7	0.000
7	USSR	152	(0.2%)	152		153.6	0.000
8	PARTIO	198	(0.2%)	245		149.9	0.000
9	USONO	181	(0.2%)	229		132.3	0.000
10	PACO	144	(0.2%)	168		118.5	0.000
11	ARMILARO	116	(0.1%)	116		116.8	0.000
12	SOCIALISMA	115	(0.1%)	115		115.8	0.000
13	POPOLOJ	123	(0.1%)	135		110.0	0.000
14	KPSU	105	(0.1%)	105		105.5	0.000
15	POLITIKO	127	(0.2%)	147		105.5	0.000
16	NECESAS	142	(0.2%)	178		105.0	0.000
17	MILITO	147	(0.2%)	189		104.5	0.000
18	EKONOMIO	108	(0.1%)	114		101.5	0.000
19	SHTATOJ	110	(0.1%)	120		99.0	0.000
20	POLITIKA	116	(0.1%)	135		95.5	0.000
21	LABORO	164	(0.2%)	238		94.8	0.000
22	NUKLEAJ	96	(0.1%)	98		93.9	0.000
23	JUNULARO	136	(0.2%)	178		93.8	0.000
24	LANDOJ	183	(0.2%)	283		93.7	0.000
25	SOCIO	109	(0.1%)	123		93.6	0.000
26	EKONOMIA	109	(0.1%)	123		93.6	0.000
27	SOCIA	108	(0.1%)	123		91.6	0.000
28	EVOLUO	108	(0.1%)	128		86.5	0.000
29	FORTOJ	94	(0.1%)	101		86.1	0.000
30	EUROPO	111	(0.1%)	137		83.7	0.000
31	SOCIALISMO	79		79		79.0	0.000
32	PROGRESO	85	(0.1%)	91		78.0	0.000
33	PRODUKTADO	82		86		77.3	0.000
34	EN	1893	(2.3%)	5685	(1.8%)	76.4	0.000
35	SOVETIAJ	75		76		73.7	0.000
36	INTERESOJ	78		82		73.2	0.000
37	SEKURECO	78		82		73.2	0.000
38	PROBLEMOJ	107	(0.1%)	141		72.7	0.000
39	KONDICHOJ	80		92		66.5	0.000
40	CK	66		66		65.7	0.000
41	RAKETOJ	66		66		65.7	0.000

	SOVETA ESPERANTISTO*(Cont.)*			ESPERANTO CORPUS			
	WORD	Freq.	%	Freq.	%	X2	p
42	RILATOJ	90	(0.1%)	114		64.9	0.000
43	POPOLO	119	(0.1%)	178		64.6	0.000
44	KAMARADOJ	72		81		61.6	0.000
45	RIMEDOJ	78		93		61.5	0.000
46	NIA	275	(0.3%)	599	(0.2%)	58.1	0.000
47	ENTREPRENOJ	62		65		58.0	0.000
48	GHIA	115	(0.1%)	179		57.7	0.000
49	PRI	868	(1.1%)	2436	(0.8%)	57.1	0.000
50	EKONOMIAJ	62		67		55.8	0.000
51	X	56		56		55.5	0.000
52	LANDO	149	(0.2%)	266		55.4	0.000
53	FESTIVALO	76		97		53.9	0.000
54	KOMITATO	72		90		52.7	0.000
55	ARMILOJ	67		80		52.5	0.000
56	HOMARO	74		95		52.0	0.000
57	INTERTRAKTADO	52		52		51.4	0.000
58	LABORULOJ	50		51		48.1	0.000
59	SCIENC	50		51		48.1	0.000
60	LIKVIDO	48		48		47.3	0.000
61	VENKO	50		52		47.0	0.000
62	TASKOJ	56		64		46.5	0.000
63	CENTRA	57		67		45.5	0.000
64	SFERO	49		52		44.8	0.000
65	VETARMADO	50		54		44.7	0.000
66	MALARMADO	47		49		43.9	0.000
67	SOCIAJ	49		53		43.7	0.000
68	MILITA	51		57		43.6	0.000
69	TEKNIKA	56		67		43.5	0.000
70	AGOJ	53		61		43.5	0.000
71	IMPERIISMO	44		44		43.2	0.000
72	NOVAJ	75		109		42.6	0.000
73	KOMUNISTOJ	43		43		42.2	0.000
74	SISTEMO	65		89		40.9	0.000
75	LENIN	42		43		40.0	0.000
76	NUKLEAN	42		43		40.0	0.000
77	PLUA	44		47		39.7	0.000
78	POLITIKAJ	49		57		39.5	0.000
79	CHIUJ	232	(0.3%)	537	(0.2%)	39.2	0.000
80	SOCIALA	40		40		39.1	0.000
81	SOCIALISMAJ	42		44		38.8	0.000
82	ADMINISTRADO	39		39		38.1	0.000
83	DISVOLVO	39		39		38.1	0.000
84	PLIFIRMIGO	39		39		38.1	0.000
85	SITUACIO	62		86		38.1	0.000
86	ORGANIZOJ	61		84		37.9	0.000
87	ORGANOJ	44		49		37.5	0.000
88	BAZO	55		72		37.1	0.000
89	TIO	454	(0.5%)	1229	(0.4%)	37.0	0.000
90	REDUKTO	39		40		36.9	0.000
91	DEMANDO	64		92		36.9	0.000
92	MINACO	40		42		36.8	0.000
93	GHIAJ	62		88		36.5	0.000
94	BATALO	53		69		36.0	0.000
95	KOMSOMOLO	36		36		35.0	0.000
96	KOMUNISTA	36		36		35.0	0.000
97	PACA	37		38		34.9	0.000
98	KONSTRUADO	40		44		34.6	0.000
99	PARTIAJ	35		35		34.0	0.000
100	MILIONOJ	49		63		33.9	0.000

SIGNIFICANT GRAMMATICAL ITEMS (within the first 1000 words)

116	TUTA	126	(0.2%)	266		29.1	0.000
124	KONTRAU	145	(0.2%)	323	(0.1%)	28.0	0.000
139	DEVAS	188	(0.2%)	456	(0.1%)	26.1	0.000
146	ILIA	82		157		25.1	0.000
152	ANKAU	366	(0.4%)	1026	(0.3%)	24.0	0.000
194	CHIO	85	(0.1%)	174		21.4	0.000
238	AJNA	23		27		17.5	0.000
251	NIAJ	100	(0.1%)	232		16.4	0.000
311	FAR	27		41		13.4	0.000

	SOVETA ESPERANTISTO *(Cont.)*			ESPERANTO CORPUS			
	WORD	Freq.	%	Freq.	%	X2	p
318	TIA	81		190		12.6	0.000
347	IGHI	18		23		11.7	0.001
372	IGHAS	20		28		11.3	0.001
377	TIES	60		135		10.8	0.001
414	NI	587	(0.7%)	1910	(0.6%)	9.9	0.002
415	POR	874	(1.1%)	2925	(0.9%)	9.9	0.002
450	CHION	57		133		8.9	0.003
512	CHIUJN	55		130		8.1	0.004
704	GHIAN	20		38		5.6	0.017
808	ESTAS	1088	(1.3%)	3804	(1.2%)	5.1	0.024
814	TIUJ	172	(0.2%)	533	(0.2%)	4.9	0.026
815	ENDAS	12		19		4.9	0.027
846	DES	23		49		4.6	0.031
960	PLI	360	(0.4%)	1201	(0.4%)	4.2	0.041

Appendix 4.2 Untypical Items in the *Soveta Esperantisto*:

Here we present the final 100 items in the Wordlist comparison. The least typical
words ate listed at the bottom of the list, not the top. These items are the least
frequently present in *Soveta Esperantisto* (i.e. those items most untypical of the
Soveta Esperantisto). The first person pronoun *mi* for example occurs so infrequently
in the *S.E.* corpus that it is calculated as being very highly significantly atypical.

	SOVETA ESPERANTISTO			ESPERANTO CORPUS			
	WORD	Freq.	%	Freq.	%	X2	p
11360	MALGRANDA	2		80		18.1	0.000
11361	AS	1		72		18.1	0.000
11362	IRIS	3		88		18.2	0.000
11363	TUTE	49		348	(0.1%)	18.2	0.000
11364	JEN	37		287		18.3	0.000
11365	SI	17		178		18.3	0.000
11366	IOMETE	1		73		18.3	0.000
11367	Ô	17		179		18.5	0.000
11368	C	5		104		18.6	0.000
11369	KREDAS	6		111		18.6	0.000
11370	PAROLIS	7		118		18.7	0.000
11371	DENOVE	13		156		18.7	0.000
11372	TROVIS	5		105		18.9	0.000
11373	STARIS	5		106		19.1	0.000
11374	SHAJNIS	1		76		19.1	0.000
11375	KELKAJN	2		85		19.4	0.000
11376	ANGLA	1		77		19.4	0.000
11377	KELKAJ	26		235		19.7	0.000
11378	RO	3		95		20.0	0.000
11379	LIAN	3		95		20.0	0.000
11380	JES	10		143		20.0	0.000
11381	L	8		131		20.3	0.000
11382	FILO	1		82		20.7	0.000
11383	SCIIS	3		98		20.8	0.000
11384	CHE	91	(0.1%)	571	(0.2%)	21.1	0.000
11385	KIUN	43		336	(0.1%)	21.6	0.000
11386	VOLAS	21		218		22.0	0.000
11387	LINGVON	3		103		22.1	0.000
11388	KROM	5		119		22.4	0.000
11389	SHIA	1		90		22.8	0.000
11390	IS	3		106		22.9	0.000
11391	VOLIS	2		99		23.1	0.000
11392	ION	6		129		23.2	0.000
11393	SIN	50		383	(0.1%)	23.6	0.000
11394	TAMEN	50		390	(0.1%)	24.9	0.000
11395	TIE	63		457	(0.1%)	25.0	0.000
11396	DO	79		538	(0.2%)	25.1	0.000
11397	APUD	2		107		25.2	0.000
11398	JAM	92	(0.1%)	602	(0.2%)	25.2	0.000
11399	VENIS	16		203		25.4	0.000
11400	HAVIS	15		197		25.4	0.000
11401	KVANKAM	5		133		26.0	0.000

		SOVETA ESPERANTISTO			ESPERANTO CORPUS			
	WORD	Freq.	%	Freq.	%	X2	p	
11402	LIAJ	4		126		26.1	0.000
11403	TRE	42		359	(0.1%)	27.2	0.000
11404	AU	96	(0.1%)	634	(0.2%)	27.2	0.000
11405	IAM	7		152		27.3	0.000
11406	UNU	99	(0.1%)	651	(0.2%)	27.6	0.000
11407	Ŭ	18		229		28.6	0.000
11408	GHIN	70		518	(0.2%)	29.5	0.000
11409	KION	36		339	(0.1%)	29.8	0.000
11410	SCIAS	26		282		29.8	0.000
11411	ONI	208	(0.3%)	1178	(0.4%)	29.8	0.000
11412	TIAM	31		311		29.8	0.000
11413	VERE	11		190		30.0	0.000
11414	DEVIS	1		117		30.0	0.000
11415	MIAJ	2		126		30.2	0.000
11416	JA	29		305		31.0	0.000
11417	VIRO	1		124		31.8	0.000
11418	TIU	237	(0.3%)	1330	(0.4%)	32.4	0.000
11419	NU	2		135		32.6	0.000
11420	TRA	8		182		33.1	0.000
11421	VIAN	2		138		33.4	0.000
11422	KIE	25		296		34.3	0.000
11423	DU	39		379	(0.1%)	34.7	0.000
11424	POSTE	15		238		35.3	0.000
11425	Í	42		399	(0.1%)	35.4	0.000
11426	JE	58		503	(0.2%)	38.7	0.000
11427	RESPONDIS	2		160		39.2	0.000
11428	PRO	80		622	(0.2%)	39.2	0.000
11429	POVIS	10		220		39.2	0.000
11430	KIAM	84	(0.1%)	643	(0.2%)	39.3	0.000
11431	TUJ	7		202		40.1	0.000
11432	EBLE	21		301		41.3	0.000
11433	CHU	84	(0.1%)	659	(0.2%)	42.2	0.000
11434	VIDIS	3		190		45.0	0.000
11435	EL	188	(0.2%)	1207	(0.4%)	47.4	0.000
11436	NE	748	(0.9%)	3713	(1.2%)	47.5	0.000
11437	POST	46		483	(0.2%)	48.6	0.000
11438	VIA	11		273		51.0	0.000
11439	LINGVO	5		230		51.4	0.000
11440	SED	330	(0.4%)	1908	(0.6%)	52.4	0.000
11441	IU	11		284		53.8	0.000
11442	Å	42		491	(0.2%)	55.7	0.000
11443	VIN	4		239		55.8	0.000
11444	IOM	9		281		56.8	0.000
11445	MIAN	1		226		58.9	0.000
11446	SE	98	(0.1%)	843	(0.3%)	63.7	0.000
11447	KE	465	(0.6%)	2633	(0.8%)	66.5	0.000
11448	LIA	13		349	(0.1%)	66.9	0.000
11449	ILI	215	(0.3%)	1500	(0.5%)	73.8	0.000
11450	ESTIS	237	(0.3%)	1675	(0.5%)	85.1	0.000
11451	LIN	3		413	(0.1%)	104.0	0.000
11452	CHAR	27		621	(0.2%)	111.2	0.000
11453	KIU	147	(0.2%)	1390	(0.4%)	121.1	0.000
11454	MIN	6		531	(0.2%)	128.7	0.000
11455	MIA	5		579	(0.2%)	143.6	0.000
11456	DIRIS	15		760	(0.2%)	170.4	0.000
11457	SHI	4		687	(0.2%)	174.5	0.000
11458	VI	53		1918	(0.6%)	399.4	0.000
11459	LI	47		2701	(0.9%)	618.8	0.000
11460	MI	106	(0.1%)	4235	(1.4%)	907.6	0.000