STATISTICAL ANALYSIS OF CONYERSATIONAL ESPERANTO,
WITH DISCUSSION OF THE ACCUSATIVE*

Bruce Arne Sherwood

Taped Esperanto conversations among skilled speakers were transcribed and statistically analyzed. The frequencies of phonemes, two-phoneme sequences, and grammatical categories were obtained. Statistics on the use of compound and derived words are presented. The most interesting data deal with the use of the accusative. It is shown that spoken Esperanto is dominantly SVO, and that other constituent orders are quite rare and restricted to special constructions. There is a discussion of the sociolinguistic and language-planning consequences of the observed accusative usage.

1. Introduction

Statistical analyses of Esperanto usage have typically been based on written texts (e.g., van Themaat 1977). Sociolinguists have emphasized not only that speech can be quite different from writing, but that informal speech is in many ways the most representative of the various ways in which people speak. Sociolinguists go to great lengths to devise interview strategies which will elicit informal speech. It would be interesting to analyze informal Esperanto speech. Of general interest, there might be special effects due to native-language influence, the greater emphasis on writing and reading due to the nature of Esperanto usage, and the possibly greater conscious awareness of formal rules on the part of Esperanto speakers. More narrowly, it is interesting to see how the language is actually spoken, in possible contrast to its prescriptions.

The audio tape service of the Esperanto League for North America (Box 1129, El Cerrito CA 94530; now http://www.esperanto-usa.org) includes in its catalog of recordings not only formal speeches but also conversations at the 1972 World Esperanto Congress held in Portland, Oregon. These conversations are presumably more self-conscious than would be ideal for some purposes, but they at least provide a more natural data base than writings or formal speeches, which may simply be readings of written papers. One of the participants (Duncan Charters, private communication) reports that during the conversations headphones and special microphones were worn in order to improve the audio quality, which may have tended to make the speech more self-conscious than desirable for the present purpose. However, another participant (William Auld, private communication) feels that the conversation between him and Peter de Smedt was rather natural despite the apparatus, partly because the taped conversation was a continuation of a conversation already in progress. Only the pauses seemed awkward, due to the consciousness that one should keep going, and some slight embarrassment is noticeable at these points on the tape.

If was hoped that contrasts could be obtained for diverse native-language backgrounds. Unfortunately, among the relatively small number of taped conversations there are no native speakers of non-Indo-European languages, which is a disappointment. Nevertheless, the present analytical study may provide a framework and a stimulus for more field work to obtain samples from more diverse and more representative types of speakers. When this study was nearly completed, it was learned that the International Cultural Service (Amruŝeva 511, 41000 Zagreb, Yugoslavia) is presently processing fifty thousand words of speech recorded at recent Esperanto World Congresses, in order to obtain conversational word frequencies for designing better teaching materials. As they are using a computer for this analysis, perhaps they will extend their work to other aspects of these conversations.

2. Methodology

The conversational material was transcribed and typed into a computer file for statistical processing. In order to facilitate this processing, the first letters of morphemes internal to words were indicated by capital letters, so that *nekredeble* 'unbelievably', for example, was entered in the form *neKredEble*. Note that grammatical endings were nor indicated unless they were internal to a word, as in *unuAFoje* 'for the first time' or *siNTeno* 'attitude', since it is

possible to identify final grammatical endings by computer program (Sherwood 1981a, 1981b). Also to facilitate computational processing participle endings were indicated by capitalization, as in *neKredAnto* 'unbeliever' or *korektIta* 'corrected'.

The frequencies of compound words, grammatical categories, and phonemes were extracted by computer program. Statistics on the order of major constituents (subject, verb, object) were determined by hand, as it appeared too difficult to do this by program. In Esperanto, the dominant order is SVO but, at least in writing, other orders are quite common, and it is of interest to have some measure of this property for conversations. Although the direct object is marked by the accusative ending "o", there are some other uses of the accusative case, which makes it difficult to identify constituent order by computer program. Another class of items which was treated by hand were inflected words (such as a verbal root used as a noun).

Compound words for the purposes of this treatment were defined simply as words which contained more than one morpheme, not counting grammatical endings. Included were not only the relatively few compounds composed of full roots such as *1aborRitmo* 'work-rhythm' but also the many compounds made by the highly productive roots traditionally called "affixes", such as *junUloj* 'young people' and *movAdo* 'movement', and even the words formed with participle endings, such as *kredAnto*. The rationale for considering *junUloj* as a compound is that *ulo* by itself is a person, and since *ul* is as free a morpheme as *jun* 'young' it seems inappropriate to call *ul* a suffix, although that is what it has traditionally been called. The example of *movAdo* 'movement' is more problematical, in that it may seem more appropriate to analyze the word as having *ad* 'continued action' as a suffix which modifies 'move', rather than to think of the word in terms of "move-like continued action".

In individual cases the distinction between compounding and suffixation seems to involve somewhat subjective judgement, within the system of Esperanto word-formation. It seems preferable to call all multi-root words "compounds", especially since the so-called "affixes" are morphemes which are as free as any other content morphemes. (By "content" morphemes I mean those which requite grammatical endings, unlike the "function" morphemes such as prepositions which do not require grammatical endings.) As for participle endings, despite the fact that *anto* or *into* 'present or past actor' are almost never used by themselves in Esperanto, such usage is not "ungrammatical" and is similar to the use of "ism" in English. In any case, all "compound" words are listed in Appendix 4 so that the reader may re-analyze them as desired.

In this analysis "inflected" words are simply those whose grammatical endings are different from the basic grammatical category of the root, as determined by checking in a standard dictionary (Wells 1969). For example, Wells lists the adjective *simila* 'similar' as the base form, so the verb *similas* 'is similar to' is counted as an inflected word. In the introduction to his dictionary Wells (1969:7-9) gives a good summary of the traditional theory that each Esperanto content root has an inherent grammatical category. Szerdahelyi (1976, 1978) has criticized this theory, arguing instead that full words are borrowed from national languages, from which a category-less root is formed by autonomous Esperanto processes. It appears that both theories lead to similar practical results. There is at least some psychological reality to the notion of inherent category, in that I nearly always found I had guessed the category correctly when I checked doubtful cases in the dictionary. A likely cause is that the Esperanto roots, coming mainly from European languages, have overtones of grammatical category that speakers of European languages will usually find natural. Another cause may be that usage patterns within Esperanto itself have over the years confirmed the category assignments of its roots, even for non-European speakers. That is, *simila* is a commonly-used adjective and is more common than *similas*, whereas *movas* 'moves' is much more common than *mova* 'motional'.

3. Data and results

The Scot William Auld, one of the most outstanding poets writing in Esperanto as well as a perceptive essayist on literary matters, and the Flemish Peter de Smedt were recorded in an interesting conversation dealing with the Hungarian Esperanto poet and novelist Julio Baghy, with the nature of translations of poetry into Esperanto, and with the need for definitive historical studies of the Esperanto literature. In a thirty-minute conversation each man spoke for almost exactly fifteen minutes. Auld spoke 1782 words (119 words per minute) end de Smedt 1948 words

(130 words per minute). In no case did there seem to be any ambiguity as to how to break up the speech stream into "words", basically because of the invariant penultimate stress and the distinctive grammatical endings.

Appendix 1 gives the phoneme frequencies for this conversation, Appendix 2 gives the frequencies of two-phoneme sequences, and Appendix 3 gives the frequencies of the various grammatical categories. These data are presented mainly because they were easily obtainable, but they can be of use in certain applications. For example, the frequencies of two-phoneme sequences have been used to plan the construction of a diphone library for purposes of speech synthesis in our laboratory.

In Appendix 4 are listed the compound words, including participles and words made with productive "affixes". Estimating the standard deviation as the square root of the number of compound words found, (7.6±0.7)% of Auld's words were compounds (136 out of 1782), and (8.9±0.7)% of de Smedt's (173 out of 1948), which are equal frequencies within the estimated errors. It is noteworthy that most of these compound words are themselves common, being encountered frequently in texts and often even as specific entries after the root in large dictionaries. There are very few really novel coinages made on the spot during the conversation. Rather, the two men pretty much limited themselves to compounds that have already been used extensively in the language. I estimate that only about three percent of the compounds are novel, in that about that many words struck me as fresh and unusual.

Of Auld's 1782 words, 211 or (ll.8±0.8)% were inflected words (e.g., a verbal root forming a noun), including 6 nouns made with "adjectival" participle endings, such as *kredAnto* 'believer'. Of these 211 inflected words, 50 were compound (multi-root) words found in Appendix 4, such as *florEce* 'in a flowery way'. For de Smedt, out of 1948 words, 232 or (11.9±0.8)% were inflected (of which 8 were participial nouns), with 60 of these appearing among the compounds listed in Appendix 4. The frequencies are:

|  | Auld | de Smedt |
|---|---|---|
| inflected only | 9.0% | 8.8% |
| compounded only | 4.8% | 5.8% |
| inflected compounds | 2.8% | 3.1% |
| totals | 16.6% | 17.7% |

All the inflected words are shown in Appendix 5. Like the compounds, the inflected words for the most part are forms which have often been seen before. Only a few percent of the words were novel coinages, though at one point in the conversation the two men were themselves amused at de Smedt's invention of the word 'in a paper-like way'.

The most interesting data concern constituent order, usually called word order, the latter being a somewhat misleading term for describing the sequence in the sentence of the major constituents subject, verb, and object. There were 179 main clauses where a transitive verb (either finite or infinitive) referred to a direct object. Despite the potential freedom of constituent order in Esperanto, which is often exploited in writing, there were only 15 examples out of 179 in which the order was not subject-verb-object (SVO; VO in the case of infinitives). Moreover, of these 15 cases, eight (examples 1-7 and 10 below) involved merely moving the demonstrative object pronoun *tion*, of which seven were of the OSV form with *tion* followed by a personal pronoun followed by a verb, and the other (example 7) was SOV, also with a personal pronoun subject (*vi tion faris* 'you that did'). Seven of these eight *tion* examples were due to Auld.

Four examples (11-14 below) involved a personal object pronoun or the common word *multon* 'much' preceding an infinitive, all due to de Smedt, who also moved the simple object *nenion* 'nothing' before a verb (15). This leaves only two examples of constituent orders other than SVO involving uncommon nouns (examples 8 and 9). It appears that these two speakers have developed individual non-SVO patterns which they use for special emphasis, but that they very rarely venture outside the confines of these stereotyped patterns. For completeness, it should he mentioned that there was one question involving (normal) OSV order: *Kion vi opinias?* 'What do you think?'. Here is a complete listing of all 15 non-SVO examples:

<center>Auld</center>

(1) *tion oni povas bedaŭr*i 'one can regret that'
(2) *tion oni faras* 'one does that'
(3) *tion mi jam diris* 'I already said that'
(4) *ne tion ni bezonas* 'we don't need that'
(5) *tion oni faras* 'one does that'
(6) *tion oni povas fari* 'one can do that'
(7) *vi tion faris* 'you did that'
(8) *min emocias, min inspiras ĉiuj aspiroj* 'all aspirations move me, inspire me';
    said with great deliberateness and emphasis
(9) *mi bibliografion faras* 'I am making a bibliography'

<center>de Smedt</center>

(l0) *tion mi ne diris* 'I didn't say that'
(11) *ĝin fari* 'to do it'
(12) *sin defendi* 'to defend oneself'
(13) *nin ĝojigi* 'to make us joyful'
(14) *multon lerni* 'to learn much'
(15) *mi nenion plu aŭdis* 'I heard nothing more'

It is striking that except for *aspiroj* 'aspirations' and *bibliografion* 'bibliography', the only words which have been moved out of SVO order are very common words. It should also be pointed out that among the 179 main clauses involving transitive verbs there are three examples where the object is the title of a book, without accusative endings, and in all of these examples the order is SVO. This latter observation may not have great significance, in that even in formal writing one sees considerable vacillation between declining or not declining proper nouns and book titles.

Another interesting aspect of this conversation is that while de Smedt six times exploited the freedom of constituent order which comes from the explicit accusative marker *-n*, he made errors just as often in his use of this marker. (Auld made no case errors.) On two occasions de Smedt failed to add the *-n* to a direct object and four times he attached an *-n* unnecessarily. In sentences 16 end 17 shown below, the *-n* has been dropped in simple SVO situations, perhaps due to the heavy dominance of SVO order which makes the explicit accusative marking rather redundant. In sentence 18, the complement of the intransitive verb "become" has incorrectly been given the accusative endings, probably due to the superficial resemblance to an SVO situation. The explanation for case errors in sentences 19 and 20 may simply be that the speaker started out intending to make an object-initial construction but got sidetracked by the long subordinate clauses, so that the noun eventually played the role of subject instead of object. In a somewhat similar manner, it is likely that in sentence 21 the speaker started out to say something like "which we like" and changed in midstream to "which can make us happy". The case errors shown in sentences 16, 17. and 18 may indicate imperfect command of the accusative in a non-native language, whereas the errors of 19, 20, and 21 appear more like typical hesitation phenomena seen even among native speakers of any language.

(16) *...eble mi trovas tricent eroj(n)...* (...possibly I will find three hundred items...)
(17) *...pretigas, jes, centpaĝan provkajeron, kiu enhavos do parto(n)...*
    (...prepares, yes, a hundred-page trial booklet, which will contain therefore a part...)
(18) *...ĝi nepre fariĝus iu<u>n</u> tre subjektiva<u>n</u> literaturhistorio<u>n</u>...*
    (...it necessarily would become a very subjective literary history ...)
(19) *Mi tamen konstatis ke ĝuste la poemo<u>n</u> kiun vi citis en via artikoleto pri tiu studo pri Hector Vermojten ĝuste temis...*
    (I nevertheless realized that precisely the poem which you cited in your article about that study about Hector Vermojten - precisely had to do with...)
(20) *0 jes, kaj iu<u>n</u> parto<u>n</u> el tiu enciklopedio, kiu estas treege bezonata, tiu estas ekzemple la historio de...* (Oh yes, and one part of that encyclopedia, which is greatly needed, that is for example the history of...)

(21) *Mi pensas ke tiu tamen estas fakto kiun povas nin ĝojigi.*
    (I think that that nevertheless is a fact which can make us happy.)

A partial analysis was made of one other conversation from the Portland conference to check in particular some of the results on constituent order. This conversation consisted of six-and-a-half minutes between the Pole Ada Fighiera-Sikorska and the Bulgarian Stojan Djoudjeff, followed by a conversation between Djoudjeff and the Briton Duncan Charters, of which eleven-and-a-half minutes were analyzed. Out of 84 main clauses involving transitive verbs, 80 (95%) had SVO order. The other four examples were these:

(22) *...mi ankaŭ tion esperas* 'I also hope that' (Fighiera-Sikorska)
(23) *...ankaŭ tiun trajton de la kongreso mi rimarkis*
    'I also noticed that characteristic of the congress' (Charters)
(24) *...la verkado ilin kaptas* 'Writing captivates them' (Djoudjeff)
(25) *...oni povas tion atingi* 'One can achieve that' (Charters)

We see the same features noted in the Auld-de Smedt data. The SVO order is highly dominant, and three of the four non-SVO orders involve moving *tion* or the pronoun *ilin*. No errors in the use of the accusative were noted for these very skilled speakers, with the possible exception of a sentence by Fighiera-Sikorska: *Mi vizitis krome Tokio(n)* 'In addition I visited Tokyo'. As mentioned earlier, there is some variation in the declining of proper nouns, although one would normally use the accusative ease in this situation, because the city name has been completely Esperantized and ends in a typical *-o*.

4. The problem of the accusative

It is well known that the accusative is the biggest source of errors for beginning students of Esperanto, and there is some awareness that even skilled speakers occasionally make case errors. This seems to be rather independent of the person's native language and occurs even with speakers as experienced as de Smedt, who is a serious student of Esperanto literature. It would seem that for most speakers the fact that the accusative poses serious problems might be explained as follows. In speaking, the SVO order is so massively dominant (92% in the Auld-de Smedt data and 95% in the other conversation, with the bulk of the remainder quite special) that it is natural to drop the *-n* as redundant on the object, since order alone normally identifies the subject and object. Although I don't have firm data, I have the impression from my own speech and from casual observation of other speakers that *-n* is sometimes added to subjects and in prepositional phrases as a form of hypercorrection, out of the guilty knowledge that one is apt to forget *-n* on objects.

The Esperanto accusative has often been attacked as being excess baggage and inappropriate in a language intended to be easy to learn and use. Apologists have countered these attacks with various arguments. One claim is that the free order benefits native speakers of languages which don't have SVO order, making it possible for them to use the constituent order which they find most natural. However, there is no real evidence that this in fact occurs: although formal studies have not been done for spoken Esperanto, in my own conversational experience I have noted that SVO is very much dominant for Hungarian and Japanese speakers (whose native order is not SVO). Moreover, there are very few languages in the world which normally place object before subject, so it is natural for all to take the second noun phrase as the object, even without an accusative marking. Sometimes contrived sentences are exhibited which show how the accusative can remove ambiguities, ignoring the fact that speakers can and do use other available constructions to avoid such problems. An example of this type of argument involves the pair of sentences *Mi trovis la vinon bonan* 'I found the good wine' and *Mi trovis la vinon bona* 'I found the wine to be good'. One can artificially construct many examples of this kind, but in practice speakers use other structures to disambiguate, including context.

Perhaps the most sophisticated defense is the one which points out that in communication between speakers from different cultural backgrounds, extra precision and redundancy are needed to compensate for the lack of shared assumptions end backgrounds. But if further studies confirm what has been found here, all Esperanto speakers share a highly dominant SVO order and a common tendency not to use the accusative correctly, which weakens the

argument. It can be agreed, however, that the accusative has great value in writing, at least optionally, because sentence structures can be much more complex than in speech. For example, active OVS sentences with complicated subjects in Esperanto often must be translated as passives into English, due to the rigid English constituent order: *La libron verkis juna fizikisto kaj sperta kemiisto* 'The book was written by a young physicist and an experienced chemist'.

It is probably fair to say that the spirited defense given for the clearly problem-ridden accusative really springs from social and political aspects of Esperanto language planning. Many historians of constructed languages have concluded that the 1905 social contract on the "untouchability" of the basic core of Esperanto was crucial in combining the necessary stability of the language with adequate capacity for evolution (Drezen 1931; Janton 1973; Golden 1977). Other constructed languages which lacked such a contract among the users tended to break up into dialects as reformers tinkered incessantly, seeking the holy grail of perfection, while the shelter of the principle of untouchability allowed the steady growth of a community of Esperanto speakers and a vital literature. A striking example is Ido, the 1907 offspring of Esperanto which was intended to remedy perceived failings of its parent, including the mandatory accusative. In Ido, the accusative ending was optional unless the object preceded the subject (Carlevaro 1978), a rule which can be observed to hold in practice for many people's spoken Esperanto. While this change was surely in the right direction on narrowly linguistic grounds, this and other changes opened up Pandora's box, leading to rapid instability as more and more of Ido was perceived to need "improvement". The leaders of the Ido movement soon found themselves having to impose artificially a period of no change, after having mocked the Esperanto "untouchability" as being a mere superstitious fetish.

It is with this historical background that justifications of the Esperanto accusative must be viewed. Speakers are naturally reluctant to discuss possible changes in the core of the language. However, the price paid in this case for stability is rather high. Insistence on "correct" use of the accusative makes learning the language more difficult. Many, perhaps most, speakers are unable to eliminate case errors even after years of experience. For speakers of moderate skill, failure to use the accusative "correctly" can lead to self-consciousness and to condescension from those more skilled. These are social effects which are undesirable in an auxiliary language. intended for easy use by ordinary people.

There is a possible resolution of the dilemma of how to remove the problem without violating the important principle of untouchability. In 1913 the forerunner of the Academy of Esperanto accepted the "principle of necessity and sufficiency" (Cherpillod 1979) proposed by René de Saussure, brother of the famous linguist Ferdinand de Saussure. This principle states that good style dictates using only those affixes really needed to fully define a word, and none that are obviated by the surrounding context. Thus *desegnaĵo* 'thing which is drawn' and *desegnado* 'continued act of drawing' can be and should he shortened to *desegno* whenever the context makes clear whether an object or an action is being described.

Perhaps the Esperanto-speaking community would be willing to admit openly that this principle is being applied by speakers to the accusative. When one says *Mi vidas kato* instead of *Mi vidas katon* 'I see a cat', the principle of necessity and sufficiency is enough to indicate that it is the cat which is seen, given that 1) SVO order is strikingly dominant; 2) the vast majority of non-SVO orders in speech involve rather special, stereotyped forms; and 3) regarding this as an "error" flies in the face of the fact that many skilled speakers do say this (and they say it precisely because there is no real ambiguity).

It is interesting that a principle of necessity and sufficiency can be seen to control the accusative in terms of semantics. Because books don't see, if I say *Libro mi vidas* or *Libro vidas mi*, there is no doubt that it is the book that is seen, despite coming first in the sentence. Richard Wood kindly pointed out to me that in the carefully edited and proofread "Der Esperantist", published in East Germany, there occurred the following "error" (Thomas 1980): "*La posttagmeza program je 16:30 h. estas aŭdebla...kaj la vespera program je 22:30 h. ni povas aŭdi...*" (The afternoon program at 16:30 is audible...and the evening program at 22:30 we can hear; should be *vesperan programon*). It seems likely that this "error" on the part of author, editor, and proofreader (all speakers of German, which has an accusative) reflects their natural application of the principle of sufficiency. Semantically, only

"evening program" can be the object in this OSV sentence, and this is the probable cause of the missing accusative ending.

Wood has sent me a number of other examples of accusative errors observed in written Esperanto. Of particular interest are two OVS sentences in which the accusative ending is missing despite the inverted order. The first is from a book review, and the second from an announcement of a conference:

(26) *Estas min instrue vidi, kio pensas profesia lingvisto...* It is now instructive to see what a professional linguist thinks... (should be *kion*)
(27) *Tiu ĉi renkontiĝo organizas Esperanto-sekcio...* This meeting is organized by an Esperanto-section... (should be *renkontiĝon*)

Again, because "professional linguist" and "Esperanto-section" are semantically the only possible subjects, the accusative ending was omitted on the object despite the fact that the object actually precedes the subject.

David Gold (personal communication) has noted that even in writing a common error is to use the accusative in a nominative slot. This may be a kind of hyper-correction. I have noticed in letters I write and receive that both extra and missing accusative markers are common, and that typists often go back and pen in an *-n* or blot one out.

An argument against extending the principle of necessity and sufficiency is that if carried to extremes one might argue that all redundancy be omitted, including the use of the grammatical endings. However, it is an observational fact that Esperanto speakers do not omit grammatical endings, nor do they seem to have much trouble with adjective-noun number agreement (no agreement errors were noted in the Auld-de Smedt conversation). In these areas description matches prescription. It is mainly in the use of the accusative that practice diverges significantly from theory.

5. Conclusions

Before statistically analyzing an Esperanto conversation it was thought that native-language influences might show up not just in phonetics but in syntax or word-formation. Such influences were not identified. It was found that constituent order is almost exclusively SVO, that other orders involve the movement almost solely of very simple objects, and that errors in the use of the accusative may be about as common as non-SVO orders, even among very skilled speakers. Although no firm data are available, my conversational experience with Japanese and Hungarian speakers indicates that these results hold true for non-SVO native languages as well, so these effects are apparently independent of native language.

The difficulties of the accusative lead to undesirable discriminations between skilled and less-skilled speakers, no matter what their native language. It might be helpful if the Academy of Esperanto would acknowledge the fact that the accusative is observed to be optional. One way to do this would be through an extension of the principle of necessity and sufficiency. The reformist problems Ido and other constructed languages encountered are probably avoidable in modern Esperanto if one speaks only of extending the existing principle in a specific, narrow area. It may be that the community will recognize that its maturity and size permit it more safe leeway than was possible 75 years ago. On the other hand, de Saussure's principle was considered just a minor stylistic interpretation of Esperanto usage, and extending this principle to something as fundamental as the accusative would be considered a big leap by many speakers. It is possible that the perils of making a fundamental change in the core of the language might outweigh the pedagogical and social benefits of simplifying accusative usage *de jure* as well as *de facto*.

NOTES

deeper association with the Department of Linguistics. This paper was presented at the 1980 Symposium on Native Language Influence on Esperanto, held in Urbana.

REFERENCES

CARLEVARO. Tazio. 1976. Eseo pri la planlingvo Ido. La Chaux-de-Fonds: Kultura Centro Esperantista.

CHERPILLOD. Andre. 1979. Neceso kaj sufiĉo. Heroldo de Esperanto, April 3.

DREZEN. E. 1931. Historio de 1a mondolingvo. Reprinted 1967, Oosaka: Pirate.

GOLDEN, Bernard. 1977. Political factors in the international language Eco-Logos 23:86.13-16.

JANTON, Pierre. 1973. L'esperanto. ("Que sais- je" series.) Paris: Presses Universitaires de France.

SHERWOOD, Bruce. 1981a. Computer processing of Esperanto text. Studies in Language Learning 3:1.145-155.

SHERWOOD. Bruce. 1981b. Komputila traktado de Esperanta teksto. In C. Bertin. ed., Unua Simpozio pri Komputiko. pp. 21-41. Rennes: the editor.

SZERDAHELYI, Istvan. 1976. la semantika modelo de Esperanto. In Zlatko Tišljar, ed.: lnternacia Lingvistika Simpozio, pp. 85-152. Zagreb: lnternacia Kultura Servo.

SZERDANELYI. Istvan. 1978. Vorto kaj vortelemento en Esperanto. Literatura Foiro no. 50. 8-10.

THOMAS, Paul. 1980. Esperanto en Radio Varsovio. Der Esperantist no. 99 (16:l.5).

VAN THEMAAT, W. A. Verloren. 1977. Productive word-formation in four natural and two constructed languages. Eco-logos 23:86.3-12.

WELLS, John. 1969. Esperanto Dictionary (in the series "Teach Yourself Books"). London: Hodder and Stoughton. New York: David &Kay.

Appendix 1
Phoneme Frequencies

|  | Auld % | de Smedt % |
|---|---|---|
| a | 11.5 | 11.2 |
| b | 1.0 | 0.7 |
| c | 0.8 | 1.0 |
| ĉ | 1.0 | 0.8 |
| d | 2.7 | 3.2 |
| e | 10.9 | 11.9 |
| f | 0.9 | 0.9 |
| g | 1.2 | 0.9 |
| ĝ | 0.6 | 0.5 |
| h | 0.9 | 0.6 |
| ĥ | 0.00 | 0.01 |
| i | 10.0 | 9.8 |
| j | 2.7 | 2.9 |
| ĵ | 0.04 | 0.09 |
| k | 4.4 | 4.6 |
| l | 5.2 | 5.1 |
| m | 3.5 | 3.0 |
| n | 7.6 | 7.2 |
| o | 8.2 | 7.4 |
| p | 3.0 | 2.6 |
| r | 5.3 | 5.8 |
| s | 7.0 | 6.7 |
| ŝ | 0.25 | 0.06 |
| t | 5.4 | 5.9 |
| u | 3.1 | 3.8 |
| ŭ | 0.3 | 0.6 |
| v | 2.2 | 2.3 |
| z | 0.3 | 0.3 |

Combined Frequencies for Both Speakers (%)

| | | | | |
|---|---|---|---|---|
| e | 11.4 | | j | 2.8 |
| a | 11.3 | | v | 2.3 |
| i | 9.9 | | g | 1.0 |
| o | 7.8 | | f | 0.9 |
| n | 7.4 | | c | 0.9 |
| s | 6.8 | | ĉ | 0.9 |
| t | 5.7 | | b | 0.8 |
| r | 5.6 | | h | 0.7 |
| l | 5.1 | | ĝ | 0.5 |
| k | 4.5 | | ŭ | 0.5 |
| u | 3.5 | | z | 0.3 |
| m | 3.2 | | ŝ | 0.15 |
| d | 3.0 | | ĵ | 0.07 |
| p | 2.8 | | ĥ | 0.005 |

Number of diphones per 3000 phonemes in a 30-minute sample of 17120 phonemes. On the left is the first phoneme, across the top the second. For example, there were 18 instances of the sequence "ed" and 27 instances of the sequence "de" per 3000 phonemes. P stands for Pause, W for word boundary. Blank indicates no occurences, 0 indicates less than 0.5. The sample contains a few proper names and foreign words. The table is continued on the following page.

|   | P | W | a | b | c | ĉ | d | e | f | g | ĝ | h | ĥ | i | j |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P |   |   | 4 | 0 | 0 | 13 | 4 | 5 | 1 |   | 1 | 0 |   | 1 | 5 |
| W |   |   | 33 | 9 | 3 | 8 | 34 | 51 | 14 | 3 | 11 | 15 |   | 17 | 8 |
| a | 9 | 60 |   | 4 | 2 | 1 | 6 |   | 4 | 3 | 1 |   |   | 0 | 39 |
| b |   | 1 | 3 |   |   |   |   | 3 |   |   |   |   |   | 1 | 0 |
| c | 1 |   | 1 |   |   |   |   | 8 |   |   |   | 0 |   | 11 |   |
| ĉ | 0 | 1 | 8 |   |   |   |   | 2 |   |   |   |   |   | 6 |   |
| d | 2 | 8 | 9 |   |   |   |   | 27 |   |   |   | 0 |   | 13 |   |
| e | 15 | 86 | 1 | 6 | 5 | 1 | 18 | 4 | 1 | 7 |   | 0 |   | 0 | 1 |
| f | 0 |   | 11 |   |   |   |   | 5 |   | 0 |   |   |   | 3 |   |
| g |   |   | 6 |   |   |   | 0 | 4 |   |   |   | 2 |   | 7 |   |
| ĝ |   |   | 1 |   |   |   |   | 1 |   |   |   |   |   | 8 |   |
| h |   |   | 7 |   |   |   |   | 1 |   |   |   |   |   | 5 |   |
| ĥ |   |   | 0 |   |   |   |   |   |   |   |   |   |   |   |   |
| i | 7 | 88 | 23 | 3 | 3 | 1 | 4 | 9 | 1 | 7 | 2 |   |   | 1 |   |
| j | 8 | 45 | 5 |   |   |   |   | 8 |   |   |   |   |   | 1 |   |
| ĵ |   |   | 0 |   |   |   |   |   |   |   |   |   |   |   |   |
| k | 0 | 3 | 36 |   | 1 |   |   | 21 |   |   |   |   |   | 16 |   |
| l | 1 | 17 | 52 | 0 |   | 0 | 4 | 16 | 1 | 1 | 0 | 0 |   | 30 | 0 |
| m |   | 9 | 10 | 0 |   |   |   | 10 | 0 |   |   |   |   | 38 |   |
| n | 17 | 49 | 10 |   | 5 |   | 9 | 29 | 1 | 6 | 0 | 1 |   | 25 | 0 |
| o | 28 | 40 | 0 | 1 | 1 |   | 2 | 3 | 1 | 2 |   |   |   | 0 | 23 |
| p |   | 7 |   |   |   |   |   | 16 | 0 |   |   |   |   | 5 |   |
| r | 1 | 13 | 31 | 0 | 0 | 1 | 3 | 32 | 1 | 0 | 0 | 1 | 0 | 29 |   |
| s | 28 | 73 | 7 |   | 5 |   | 0 | 23 |   |   |   |   |   | 8 |   |
| ŝ |   |   | 3 |   |   |   |   |   |   |   |   |   |   | 1 |   |
| t | 0 | 1 | 39 |   |   |   |   | 31 | 0 |   |   | 0 |   | 37 |   |
| u | 4 | 30 | 3 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |   |   | 2 | 4 |
| ŭ | 1 | 10 |   |   |   |   | 1 |   |   | 0 | 1 |   |   |   |   |
| v |   |   | 16 |   |   |   |   | 15 |   |   |   |   |   | 21 |   |
| z |   | 1 | 1 |   |   |   |   | 3 |   |   |   |   |   | 2 |   |

| | ĵ | k | l | m | n | o | p | r | s | ŝ | t | u | ŭ | v | z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P | 0 | 40 | 5 | 12 | 5 | 2 | 3 | 0 | 12 | 1 | 4 | 1 | | 2 | |
| W | 0 | 39 | 53 | 53 | 32 | 12 | 46 | 12 | 20 | 3 | 48 | 4 | | 28 | 0 |
| a | 1 | 6 | 21 | 13 | 30 | | 4 | 23 | 64 | | 13 | 1 | 13 | 8 | 1 |
| b | | | 7 | | | 6 | | 1 | 1 | | | 1 | | | |
| c | | 1 | | | | 3 | | | | | 0 | | | | 0 |
| ĉ | | | | | | 1 | | | | | | 8 | | | |
| d | | | | 0 | | 17 | | 2 | | | | 6 | | | |
| e | | 10 | 16 | 10 | 45 | 1 | 2 | 42 | 39 | | 7 | | | 7 | 3 |
| f | | | 1 | | | 4 | | 1 | | | 0 | 0 | | 0 | |
| g | | | 2 | | 1 | 2 | | 3 | | | | 1 | | 2 | |
| ĝ | | | | | | 2 | | | | | | 4 | | | |
| h | | | 0 | | | 5 | | | | | | 1 | | | |
| ĥ | | | | | | | | | | | | | | | |
| i | | 8 | 6 | 5 | 22 | 23 | 1 | 7 | 28 | | 11 | 25 | | 3 | |
| j | | | | | 9 | 1 | | | | | 1 | 2 | | | |
| ĵ | | | | | | 1 | | | | | | 0 | | | |
| k | | | 3 | | | 24 | | 5 | 2 | | 7 | 5 | | 2 | 2 |
| l | | 2 | 1 | 1 | 1 | 11 | 1 | | 0 | | 3 | 6 | | 0 | |
| m | | 0 | | | | 14 | 10 | | | | | 2 | | | |
| n | | 7 | 1 | 0 | | 7 | | 0 | 11 | 0 | 21 | 11 | | 1 | 1 |
| o | | 1 | 12 | 15 | 50 | | 7 | 18 | 8 | | 0 | | | 11 | 1 |
| p | | | 12 | | | 17 | | 21 | | | 1 | 1 | | | |
| r | | 8 | 3 | 2 | 1 | 26 | 0 | 0 | 2 | 0 | 4 | 2 | | 0 | |
| s | | 2 | 0 | 0 | | 5 | 6 | | 0 | | 36 | 4 | | 0 | |
| ŝ | | | 0 | | | | | | | | 1 | | | | |
| t | | 0 | | 0 | | 22 | 0 | 17 | 1 | | 0 | 13 | | | |
| u | | 5 | 4 | 1 | 16 | 2 | 1 | 9 | 8 | 0 | 5 | | | | 1 |
| ŭ | | | | | | | | 1 | | | 1 | | | 0 | |
| v | | 0 | | | | 9 | | 0 | | | | 4 | | | |
| z | | | | | | 2 | | | | | | | | | |

## Appendix 3
## Grammatical Categories

| | Auld | | de Smedt | |
|---|---|---|---|---|
| adjective | 6.8% | (11.5) | 8.0% | (15.5) |
| noun | 13.7% | (24.9) | 14.2% | (23.5) |
| pronoun | 12.0% | ( 9.8) | 10.3% | ( 6.5) |
| correlative | 7.5% | (27.6) | 7.6% | (17.4) |
| adverb | 6.3% | ( 1.8) | 6.6% | ( 0.8) |
| past | 2.9% | | 3.6% | |
| present | 10.3% | | 8.4% | |
| future | 0.7% | | 0.7% | |
| infinitive | 1.9% | | 2.7% | |
| imperative | 0.4% | | 0.6% | |
| conditional | 0.4% | | 0.8% | |
| other | 37.1% | | 36.6% | |

The percentages in parentheses represent the fraction of each category that was in the aecusative case; For example, of Auld's 1782 words, 122 or 6.8% of them were adjectives, and 11.5% of these 122 adjectives (14) were in the accusative case. The accusative case for adverbs indicates "direction towards".

Appendix 4
Compound Words

Auld :136 compound words out of 1782 total words = (7.6±0.7)%
(1782 words)/(l5 min.) = 119 words per minute

| | | |
|---|---|---|
| reAgis | kredAnto | falPuŝIĝo |
| verkIsto | siNTenon | pasIntFoje |
| iaSence | tolerEmaj | pretIgis |
| melodramEca | amikIĝas | tradukItaj |
| amEgas | tolerEmaj | neEble |
| ĝojIgas | treEge | neKredEble |
| klarIgi | dekSepa | neKredEble |
| iomEte | jarCento | tradukIta |
| siNTenon | pliBonIgo | finIĝos |
| surdEcon | bonEco | deNove |
| parolAnte | diKredAnto | malGranda |
| publikIg1taj | sentIgos | komprenEble |
| partEton | neKredAnto | multOblIgas |
| lingvAĵo | homArAno | pliParto |
| senEmocia | dirIte | dronIntaj |
| BaghyEcon | legAdo | dronIntaj |
| florEce | BudhAno | movAdo |
| mirEgis | homKredaj | literaturHistorio |
| aktorEca | finFine | komprenEble |
| teatrEca | homAro | iomEte |
| deNove | deVojIĝis | verŜajne |
| tradukAnto | belEga | kunLaboras |
| poemAroj | rimFormo | reVerki |
| lulKanto | kiaManiere | duFoje |
| poemAro | enTute | perfektEcon |
| foliAro | enHavo | elDonas |
| deNove | bonEge | alDoni |
| sentimentalAĉo | deNove | laborRitmo |
| pomArbo | literaturHistorio | tutSimple |
| kvinDek | interRompi | proksimUme |
| plurfoje | fidInda | eksModIĝInta |
| forPelas | kunLaborAntojn | ĉiuOkaze |
| enuIga | konsentIte | enTute |
| disVolvas | malNovajn | .elDonas |
| perdIĝas | elSerĉis | mondMilito |
| farIĝas | enIras | poKajere |
| trenIta | movAdon | bindIgas |
| junUloj | iomEte | malLongDistanca |
| duDek | unuAVice | elDonos |
| reLegis | bonŜanca | deNove |
| reLegi | duAVice | tutSimple |
| gurdIte | du-tri | naciLingvAnoj |
| recenzAnto | malGrandaj | enŜtopas |
| kvinDek | iomEte | suprAĵon |
| malBona | solvIta | reVido |
| kiaManiere | | |

de Smedt: 173 compound words out of 1948 total words = (8.9±0.7)%
(1948 words)/(l5 min.) = 130 words per minute

| | | | |
|---|---|---|---|
| treEge | reVenas | finFine | reVeni |
| legAdon | deNove | duCent | supozIgas |
| eksModaj | artikolEto | elDonIsto | korektIta |
| iomEte | diKredo | interTempe | ordIgIta |
| treEge | komprenEble | kvinDek | komprenEble |
| komparEbla | treEge | malNovRusa | neKompleta |
| malFacile | tradukInto | tutSimple | preferInde |
| treEge | tutSimple | Esperanto-libroj | reElDonu |
| romantikEco | finAĵo | malGrandaj | malNovan |
| arĥaikEco | literaturHistorio | librEtoj | deLonge |
| malAperis | preferInde | kunIgIta | havEbla |
| foliAro | literaturHistorio | ŝlosIlo | komprenEble |
| kompilAĵo | enMeti | tradukAĵo | ĝisData |
| kritikIsto | sentIĝas | oficEjo | reElDonu |
| ĉefVerko | junUloj | Esperanto-revuoj | malNovan |
| treEge | junUloj | malLongan | elDonu |
| unikAĵo | treEge | malLonga | komprenEble |
| treEge | interesIĝas | aperIgi | treEge |
| malPlaĉis | eblEcon | reSendo | Esperanto-movAdoj |
| hibridEco | elDonIstoj | ankoraŭFoja | urĝEga |
| miksAĵo | reElDonis | reSendo | Esperanto-movAdo |
| aliFlanke | reElDonis | starIgas | siNSekvo |
| treEge | reElDonita | verkItaj | malAkordIĝoj |
| recenzAnto | elDonItaj | ĉeEstis | movAdo |
| daŭrIgi | geFratoj | komitatKunVenon | disKonIgi |
| aperIgIntajn | reElDono | komitatKunVenon | eblEcon |
| farIĝus | reElDono | elDonIsto | tutSimple |
| literaturHistorion | ĝojIgi | farIĝos | forLasis |
| treEge | daŭrIgi | multEKosta | forLasis |
| recenzisIston | literaturHistorio | aperIgo | movAdon |
| verŜajne | komprenEble | interAlie | malĜusta |
| neKredAnto | kunLaborAnton | antaŭVidi | treEge |
| malFacile | kunLaborAnton | aperDaton | malFrue |
| pastro-poeto | komprenEble | aliFlanke | kunSidon |
| neKredAnto | tutSimple | ĝojIga | forIri |
| konAta | ekScii | sciIgo | komprenEble |
| ateIsto | reSpegulIgas | pretIgas | treEge |
| malAntaŭ | kunLaborAnton | pretIgas | interParoli |
| verkAnto | iomEte | centPaĝan | treEge |
| verkOnoto | komprenEble | provKajeron | reVido |
| kiteraturHistorio | malOfte | enHavos | |
| aperIntajn | triCent | malHelpo | |
| treEge | triCent | elDoni | |
| inteAlie | tradukAĵoj | elDonis | |

14

Auld

| | | | | |
|---|---|---|---|---|
| fakte | interese | aspiroj | nome | same |
| favore | aŭtuna | vero | bone | tutSimple |
| rilate | deNove | pliBonlgo | eble | proksimUme |
| angla | valora | diKredAnto | iomEte | ĉiuOkaze |
| rilatoj | absolute | veron | unuAVice | enTute |
| similas | ĝuste | kredo | duAVice | mia |
| iasence | sonĝe | inspiron | escepte | dua |
| melodramEca | plurFoje | kredo | relative | poKajere |
| ekzemple | absolute | neKredAnro | verkoj | poste |
| kore | verko | nome | pravas | kompleton |
| mia | ĝuste | kredoj | bone | relative |
| iomEte | tede | fidoj | fakte | malLongDistanca |
| mian | Esperanta | Infana | aktuale | mian |
| siNTenon | opinion | eksplicite | iomEte | deNove |
| interese | verko | dirIte | angla | bone |
| sincere | publike | temas | ĝuste | tutSimple |
| parolAnte | tute | analogia | pasintFoje | nian |
| tradukoj | trompa | religia | verkoj | pravas |
| temas | ekzemple | sento | neEble | absolute |
| pravas | Esperanta | sento | neKredEble | propagandan |
| interese | Esperantajn | absolute | neKredEble | reciproke |
| ekzemple | opinion | ideologia | absolute | stimule |
| anglan | simple | legAdo | kvante | reVido |
| angla | gurdIte | katolikaj | kvante | |
| senEmocia | absolute | sentoj | angla | |
| florEce | pravas | budhaj | deNove | |
| ĝuste | simple | sentoj | komprenEble | |
| Sanga | unua | homKredaj | verkoj | |
| anglan | recenzAnto | sentoj | forgeson | |
| absolute | opinion | finFine | kare | |
| aktorEea | Esperanta | aspirojn | rusa | |
| teatrEca | eble | koincido | gazEto | |
| melodrama | kiaManiere | koincido | nederlanda | |
| deNove | opinio | ĝuste | komprenEble | |
| temas | kredAnto | kiaManiere | iomEte | |
| tradukAnto | ofte | enTute | verŜajne | |
| verkojn | siNTenOn | nederlanda | domaĝe | |
| evidente | ĝuste | bonEge | eblon | |
| unuaj | ekzemple | deNove | kulpas | |
| nome | treEge | ekzemple | angla | |
| dua | alte | bone | fine | |
| unua | ekzemple | kunLaborAntojn | tria | |
| Siberia | signife | konsentlte | duFoje | |
| lulKanto | anglaj | fakte | celas | |
| vere | religiaj | inkluzive | verkon | |
| valoraj | dekSepa | sciencaj | ekzemple | |
| poste | ĝuste | verkoj | poste | |

| | | | | |
|---|---|---|---|---|
| treEge | papere | verko | fakte | verko |
| naciaj | treEge | ana1ogian | malOfte | aliFlanke |
| ekzemple | tute | tutSimple | eble | malHelpo |
| nederlanda | fakte | eblas | nederlanda | alfabeta |
| flandra | vere | ekzemple | finFine | komenco |
| legAdon | aliFlanke | eblas | interTempe | poste |
| verkoj | treEge | eblas | eble | komprenEble |
| fakte | nia | vere | ekzemple | bone |
| eksModaj | opiniojn | preferInde | nederlanda | preferInde |
| iomEte | opinion | tradukojn | malNovRusa | deLonge |
| liaj | unua | verkojn | jugoslava | komprenEble |
| verkoj | recenzAnto | verkoj | tute | ĝisData |
| treEge | literatura | eblas | hararde | mia |
| naciaj | unea | precipe | tutSimple | komprenEble |
| komenca | ĝuste | cetere | nacia | treEge |
| ĝuste | sufiĉe | cetere | kunIgIta | ekzemple |
| ekstreme | mian | treege | bindo | naciaj |
| fakte | opinioj | vere | hazarde | urĝa |
| tute | kredo | vere | nederlanda | movAdo |
| ĝuste | treEge | niaj | fakte | nacia |
| nia | verŜajne | sufiĉe | vere | siNSekvo |
| nederlanda | fakte | multe | vere | bataloj |
| fakte | tute | sanga | eble | kvereloj |
| fakte | neKredAnto | tradukoj | komenco | ĵaluzoj |
| malFacile | malFacile | unua | oederlanda | movAdo |
| nederlanda | nederlanda | vere | poste | tipe |
| verkon | eble | eble | mia | pravis |
| treEge | forto | fakre | unuan | sian |
| Esperantan | verkAnto | inkluzivi | fakte | opinion |
| tradukon | verkOnto | tradukojn | sendo | tutSimple |
| Esperanta | treEge | komprenEble | reSendo | oficiale |
| traduko | cetere | plene | ankoraŭFoje | movAdon |
| fakte | fakte | kunLaborAnton | reSendo | unue |
| troa | inrerAlie | kunLaborAnton | personajn | poste |
| senton | fakte | komprenEble | opiniojn | vero |
| verkoj | deNove | nederlanda | fakte | treEge |
| fakte | ĝuste | tradukoj | vere | necesas |
| verko | ĝuste | tradukoj | necesas | malFrue |
| bone | diKredo | tutSimple | cetere | kunsidon |
| aŭtuna | komprenEble | tradukoj | eble | komprenEble |
| verko | cetere | valoraj | ĝuste | treEge |
| fakte | precipe | kunLaborAnton | dana | multon |
| ege | treEge | Fakte | multEKosta | treEge |
| ĉefVerko | fakte | iomEte | verko | reVido |
| tute | traduklnto | nederlanda | verko | |
| treEge | tradukon | komprenEble | interAlie | |
| sanga | literatura | tute | raporto | |