[*Proceedings of the National Symposium on Machine Translation*, UCLA February 1960]

Session 5: GRAMMATICAL STUDIES

AUTOMATIC ENGLISH INFLECTION[1]

William D. Foust
Harvard University

This paper is a description of a system for classifying English nouns and verbs into inflectional classes semi-automatically, and for inflecting English automatically, using these inflectional classes. The inflectional classification system has been applied to the English correspondents in the Harvard automatic dictionary file, and the inflection process itself is intended to be a part of the English synthesis procedure in translation.

The inflectional classes were developed by Grant and Matejka [1]. In their system each noun paradigm has two members, singular and plural; each verb paradigm has five members, infinitive, third-person singular present, past, past participle, and present participle. There are six noun inflectional classes and eight verb classes in the system. A summary of the noun classes is given in Figure 1. For example, the plural of class 1 nouns is formed simply by adding "s" to the singular, the plural of class 2 nouns by adding "es", and the plural of class 4 nouns by changing "y" to "i" and adding "es". In verb class 1, for example, the principal parts are formed by adding "s", "ed", "ed", and "ing" to the infinitive base; in verb class 6 "es", "ed", "ed", and "ing" are added.

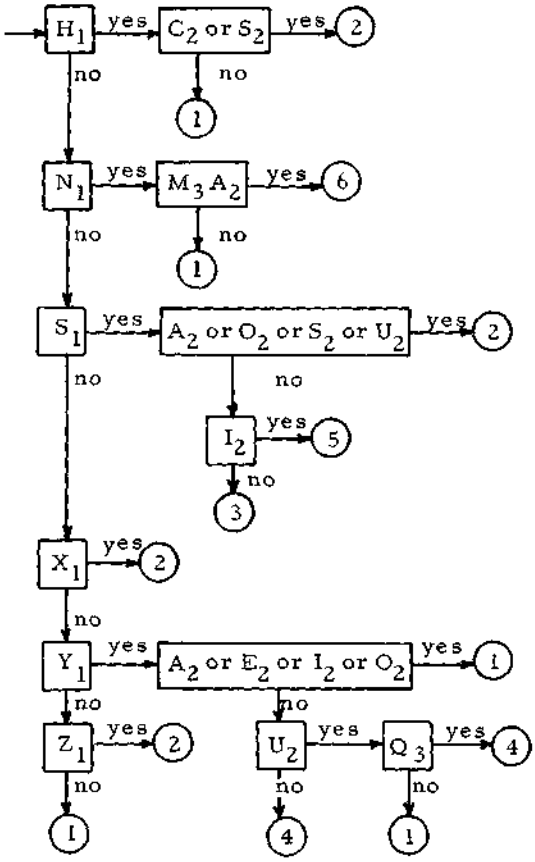
Class	Formation of Plural (from singular)	Example
1	+ s	books
2	+ es	benches
3	unchanged	species
4	y > i + es	abilities
5	is > es	syntheses
6	man > men	weathermen

Noun Inflectional Classes

Figure 1

[1] This work has been supported in part by the National Science Foundation and by the Rome Air Development Center, Air Research and Development Command.

Nearly all nouns and verbs in canonical form, that is, singular in nouns and infinitive base in verbs, can be assigned to a particular inflectional class determined by the last few letters of the canonical form. For example, as shown in Figure 2, where letters denote themselves and subscripts denote the ordinal position of the letter counting from the end of the word, a noun ending in "h" belongs to class 1 and so its plural is formed by adding "s", unless the noun



Noun Classification

Figure 2

ends in "ch" or "sh", in which case it belongs to class 2 and its plural ends in "es"; thus, "paths" and "lengths", but "benches" and "marshes".

There are exceptions which cannot be correctly classified by inspecting the final letters; for instance, "stomach" does not add "es" like most nouns in "ch", but adds "s". Other examples are "child", "half", "datum", and the verbs "stand" and "singe".

The classification program checks an exception list before applying the regular classification procedure. If the noun or verb is not an exception, the program runs through a tree branching on final letters; the tree has specific English inflectional classes for endpoints. The entire tree for nouns is shown in Figure 2.

There is one important case in which the tests do not lead to unique classification. Some verbs ending in consonant + vowel + consonant double the final consonant in certain forms, others do not. For example, the "r" in "infer" is doubled, but it is not in "suffer"; the orthography does not indicate stress, which is almost always the criterion in this case. Verbs of this type, which cannot be classified unambiguously by the program, are recorded on a separate list to be classified manually.

This classification procedure has been applied to all the noun and verb correspondents in our dictionary file. It was assumed that English correspondents of Russian nouns should be classified as nouns and likewise for verbs. One complication in the classification and inflection of English correspondents is the possibility that a single Russian word may have a correspondent consisting of several words. Some examples of these multiple-word correspondents are "take aim", "intensity of magnetization", and "steel facing". Once a marker has been stored indicating which word of the string should be classified and inflected, multiple-word correspondents can be treated like the rest.

The inflection program, which will form a part of the translation procedure, operates on each English noun or verb correspondent in the text in a manner determined by the inflectional class marker associated with the correspondent and by the interpretation of the function of the Russian word. The interpretation of the function is generated in the morphological analyzer [2] and is reduced in ambiguity in the syntactic analyzer [3] . In noun interpretations, the inflection program distinguishes only non-singular nouns from the rest. For verbs the possible distinctive interpretations are

present third-person singular, present not third-person singular, future first-person, future not first-person, infinitive, imperative, present gerund, past gerund, and the four participles. It is assumed that, as the first approximation, the correlation of number in nouns and tense in verbs is exact between Russian and English.

An English correspondent is inflected only if either it belongs to one of the six noun paradigms or one of the eight verb paradigms or else appears on a short table of paradigms of frequent irregular verbs, such as "be", "do", and "have". However, some inflected forms can be formed without knowing the class marker specifically, the infinitive, the imperative, and the future.

Suppose, for example, that a Russian form in a text is interpreted as a past gerund; all its correspondents with regular class markers will be inflected as "having" + past participle: so that, if one of the correspondents is "act" with class marker 1, the past participle is formed by adding "ed" to the canonical form, and the final inflected form is "having acted".

The system discussed here is only a first approximation to the proper inflection of English, but it will provide a substantial foundation for a system of English synthesis.

REFERENCES

[1] Grant, S., and Matejka, L., "The Automatic Classification and Inflection of English Words", Mathematical Linguistics and Automatic Translation, Report No. NSF-3, Harvard Computation Laboratory, Section XIII (1959).

[2] Sherry, M. E. , "Automatic Affix Interpretation and Reliability of the Harvard Automatic Dictionary", Proceedings of the National Symposium on Machine Translation, Los Angeles, (1960).

[3] Sherry, M. E., "Predictive Syntactic Analysis", Proceedings of the National Symposium on Machine Translation, Los Angeles (1960).