American Journal of Computational Linguistics *Microfiche 34 : 19*

A Tuneable Performance Grammar

Jane J. Robinson

Artificial Intelligence Center
Stanford Research Institute
Menlo Park, California 94025

ABSTRACT

This paper describes a tuneable performance grammar currently being developed for speech understanding. It shows how attributes of words are defined and propagated to successively larger phrases, how other attributes are acquired, how 'factors' reference them to help the parser choose among competing definitions in order to interpret the utterance correctly, and how these factors can easily be changed to adapt the grammar to other discourses and contexts. Factors that might be classified as 'syntactic' are emphasized, but the attributes they reference need not be, and seldom are, purely syntactic.

Acknowledgment

This research was supported by the Defense Advanced Research Projects Agency of the Department of Defense and monitored by the U.S. Army Research Office under Contract No. DAHC04-75-C-0006.

A performance grammar (PG) defines the form and meaning of the kinds of utterances that occur in spontaneous dialog. When the definitions of the grammar provide information that helps a parser choose those rules most likely to lead to correct interpretations of utterances, the grammar is said to be 'tuned'. When the tuning is easily changed when the domain of discourse changes, the grammar is said to be 'tuneable'. The ability to tune a grammar is particularly important in speech understanding where the inherent uncertainty of the input causes false paths through the grammar to be multiplied.

This paper describes a tuneable PG being developed jointly by SRI and SDC for a computer-based speech understanding system. Its vocabulary and phrase types, selected from protocols, are appropriate for asking and answering questions about properties of submarines. The PG now defines over 70 word and phrase categories. Its scope extends far beyond syntax. A discourse component enables it to handle anaphora and ellipsis, as in: "What is the surface displacement of the Lafayette?.... What is its draft?", and "What is the length of the Lafayette?..., The Ethan Allen?" A semantics component defines a common meaning for paraphrases, as in "the speed of the Lafayette is 30 knots" and "the Lafayette has a speed of thirty knots". (See Walker et al., 1975; Paxton and Robinson, 1975; Hendrix, 1975; Deutsch, 1975.)

Each definition composing the PG has three parts. The first names a word category or a phrase category and provides a context-free production for its composition. The second part,

called 'attributes', tells how to determine the properties of an instance of the category. Any definition can reference multiple sources of knowledge--acoustic, syntactic, semantic, discourse, or pragmatic--for information needed to determine attribute values. The third part, 'factors', defines scores for combinations of attributes, indicating how well they 'fit'. It is through factor scores that the grammar is tuned. The individual scores are combined into a composite score which is used by the parser to choose among competing parsings. A purported instance of the definition with a score of OUT for any factor is immediately eliminated; a low score may eliminate a parsing path; a high score enhances the priority of a parsing path that applies the definition.

Our mnemonic terms for factor scores are VERYGOOD, GOOD, OK, POOR, BAD, and OUT. These are estimates of likelihood. They are necessarily vague, because we are dealing with gradual phenomena and probabilistic tendencies. They mean something like "quite likely", "expected", "ordinary", "odd but possible", "unlikely--listen again", and "so special that we do not define it". Rigid, prescriptive judgments are avoided. Combining "foot" with "-s" as a plural noun is indeed wrong and therefore OUT. On the other hand, "fuel" does combine with plural "-s" with the specialized meaning "kinds of fuel". At present, "fuels", like "foots" is judged to be OUT for our language, but this judgment can easily be altered, if we find that our language users refer to kinds of fuel as "fuels".

Since factor scores can be changed without affecting the rest of the definition, the grammar is tuneable to different discourse domains and styles of speaking. Also, if one factor defines a low score for an instantiation, others may still raise the composite score. A statistically improbable phrase that makes sense and is uttered intelligibly should not be unduly difficult to recognize and interpret.

The rest of this paper examines sequences of definitions required for parsing and understanding a typical utterance. We begin with word definitions, and show how the attributes of words are propagated to successively larger phrases, how other attributes peculiar to higher-level phrases are added, and how factors reference them in tuning the grammar.

Preceding discourse and underlying semantic distinctions constrain the surface syntax of an utterance. Because superficial syntactic properties signal those constraints, it is often economical to use syntactic factors in order to disconfirm a wrong parsing path or confirm a correct one, avoiding calls on semantics, discourse, and acoustics for expensive in-depth evaluations. For example, if someone says "fuel supplies", we do not want the parser to explore in depth the application of rules that build a plural noun-phrase from "fuel s..." without considering an alternative definition in which "fuel" is a modifier of a countable nominal beginning with "s". To this end, we include a factor that checks the countableness of "fuel" by referencing a count/mass/unit (CMU) attribute, which is syntax

oriented but essentially semantics based.

Examples of some useful syntax-oriented attributes defined for the word category N (noun stem) appear in (1) below. Every N has a value for the CMU attribute drawn from the set (COUNT MASS UNIT). Ns with the CMU value UNIT (such as "foot", "ton", "knot") combine easily with plural suffixes and number expressions (e.g., "two knots", "five feet"), but not so well with definite determiners ("those two knots"), or genitive suffixes ("the twenty knots' speed"). (Cf. "the Ethan Allen's speed".)

```
(1)    WORDS,DEF      N
         FUEL                        CMU = (MASS);
         FOOT                        CMU = (UNIT), PLSUFF = NO;
         LAFAYETTE                   CMU = (COUNT);
         SURFACE,DISPLACEMENT        CMU = (COUNT), RELN = T;
         TON                         CMU = (UNIT);
```

Like the CMU attribute, the RELN attribute is essentially semantic. It marks such words as "surface displacement", "speed", "length", and "draft" as special 'relational' noun words. Syntactically, relational Ns do not combine readily with plural suffixes and number expressions, and when they do, the meaning is specialized. To some degree, they are like mass Ns; "three speeds" (three rates of speed) is analogous to "three fuels" (three kinds of fuel). However, "a speed of twenty knots" is acceptable, while "a fuel of two tons" is ill formed.

The attribute PLSUFF distinguishes irregular plurals like "foot". Unlike the CMU and RELN attributes, it is purely

syntactic.

Attributes affecting the ability to combine with the plural suffix "-s" are referenced in the two composition rules of (2), defining the category NOUN. The attribute statements propagate the attributes of the stem, adding a number attribute (NBR). The first factor of N1 references the CMU attribute and states that if the value is MASS, then the score is GOOD. This judgment incorporates our knowledge that the other rule, N2, cannot apply to mass noun-stems. If the token is a mass noun-stem, N1 is the right composition rule to apply.

```
(2)     RULE,DEF N1      NOUN = N;
            ATTRIBUTES
                CMU,RELN,PLSUFF FROM N, NBR = "(SG);
            FACTORS
                CMU = IF CMU EQ "(MASS) THEN GOOD ELSE OK,
                RELN = IF RELN EQ "T THEN GOOD ELSE OK,
                PLSUFF = IF PLSUFF EQ "NO THEN GOOD ELSE OK;
            EXAMPLES
                SURFACE DISPLACEMENT, FOOT, FUEL (GOOD)
                SUBMARINE (OK);

        RULE,DEF N2      NOUN = N -PL;
            ATTRIBUTES
                CMU,RELN,PLSUFF FROM N, NBR = "(PL);
            FACTORS
                PLSUFF = IF PLSUFF EQ "NO THEN OUT ELSE OK,
                CMU = IF CMU EQ "(MASS) THEN OUT ELSE OK,
                UNIT = IF "UNIT IN CMU THEN GOOD ELSE OK,
                RELN = IF RELN EQ "T THEN POOR ELSE OK;
            EXAMPLES
                FOOT -S, FUEL -S (OUT), TONS (GOOD)
                SURFACE DISPLACEMENTS (POOR), SUBMARINES (OK);
```

Like the CMU factors, the PLSUFF factors enhance the score for applying N1 to stems that do not take a plural suffix and constrain N2 not to apply. A RELN factor enhances the score when

N1 is applied to a relational noun stem, and lowers the score when N2 is applied. Plurals of such stems are less likely than singulars, but are possible. The UNIT factor of N2 enhances the score when the composition rule applies to an N with the value UNIT in its CMU attribute. This judgment is based on the fact that, in our current task domain, all the measured properties have measurements exceeding one unit and on the reasonable expectation that exactly one unit is a special case.

The attributes of the Ns continue to be propagated through successive composition rules so that noun phrases acquire the attributes, with some exceptions and additions, of the Ns that are their heads. One of the additional attributes of NPs is FOCUS: a noun phrase is definite (DEF) or indefinite (INDEF).

Combining definite focus with a unit is unusual. Compared with "which submarine", "which seven thousand tons" seems odd, as does "those twenty knots" and "a draft of the five feet". Indefinite focus is more common for units: "a ton", "a draft of five feet", "twenty knots". It does not suggest a uniquely determinable object or set of objects, pointed to in the discourse.

How these syntactic tendencies are handled in three composition-rule definitions is shown in (3). Factors in NP4 eliminate expressions like "five fuels", "five submerged speeds of three knots", "how much submarine", "one submarines" and "how many fuel". They score expressions like "five feet" as VERYGOOD and "five submarines" as OK. Factors for NP7 eliminate "those

submarine", "those fuels" and accept "what fuel" as OK, while "which tons" and "that draft of five feet" are POOR. Factors for NP11 eliminate "a fuel", "a draft of the Lafayette", and "a submarines"; accept "a submarine", "a ton", "the submarine", and "the submerged speed", and score "the ton" and "the draft of five feet" as POOR.

```
(3)  RULE,DEF NP4     NP = NUMBERP NOM;
      ATTRIBUTES
          FOCUS = "INDEF, MOOD,NUM FROM NUMBERP, RELN FROM NOM,
          NBR = GINTERSECT(NBR(NUMBERP),NBR(NOM)),
          CMU = GINTERSECT(CMU(NUMBERP),CMU(NOM));
      FACTORS
          CMU = IF NULL CMU THEN OUT ELSE OK,
          HUN = IF FSTWD(NUMBERP) IN "(HUNDRED THOUSAND MILLION)
              THEN OUT ELSE OK,
          NBR = IF NULL NBR THEN OUT ELSE OK,
          UNIT = IF "UNIT IN CMU THEN VERYGOOD ELSE OK,
          RELN = IF RELN EQ T THEN OUT ELSE OK;

RULE,DEF NP7     NP = DET NOM;
     ATTRIBUTES
          FOCUS = "DEF, RELN FROM NOM, MOOD FROM DET,
          CMU = GINTERSECT(CMU(DET),CMU(NOM)),
          NBR = GINTERSECT(NBR(DET),NBR(NOM));
     FACTORS
          CMU = IF NULL CMU THEN OUT ELSE OK,
          UNIT = IF "UNIT IN CMU THEN POOR ELSE OK,
          NBR - IF NULL NBR THEN OUT ELSE OK;

RULE,DEF NP11     NP = ART NOM;
     ATTRIBUTES
          RELN FROM NOM, FOCUS FROM ART, MOOD = "DEC,
          CMU = GINTERSECT(CMU(ART),CMU(NOM)),
          NBR = GINTERSECT(NBR(ART),NBR(NOM));
     FACTORS
          CMU = IF NULL CMU THEN OUT ELSE OK,
          NBR = IF NULL NBR THEN OUT ELSE OK,
          UNIT = IF "UNIT IN CMU AND FOCUS EQ "DEF
              THEN POOR ELSE GOOD,
          RELN = IF RELN EQ T AND FOCUS EQ "INDEF AND
              CMU EQ "(COUNT) THEN OUT ELSE OK;
```

In each definition, a UNIT factor references the CMU

attribute of the NP. If the value is NIL, the definition is not applicable. If UNIT is a value, then the UNIT factor for NP4 scores the application as VERYGOOD. There are two reasons for this judgment. Number expressions are typically found with unit expressions to form measure expressions, and units are more likely to occur with indefinite than with definite focus, as the preceding examples ("twenty knots" and so on) have indicated.

Since the focus for NP7 is always definite, the UNIT factor decreases the score for applying it when the UNIT value appears in the CMU attribute. For NP11, the UNIT factor scores the application GOOD if the article is "a" and UNIT appears in the CMU values, but POOR if the article is "the".

NP4 applies especially well to instances in which units are present, but does not apply at all if the head of the nominal constituent is a RELN stem. In discourse about washing machines and bicycles, "three speeds" might occur in an ordinary way, but for our current discourse, we do not anticipate such a combination. Certainly, we do not expect "three surface displacements".

Such constraints relieve the need for detailed analysis. For instance, assume that the acoustic mapper has tentatively offered both "submarine" and "submerged speed" as acoustically plausible alternatives for filling the gap in the partially analyzed phrase "three ----- -s of the U.S. Navy". This is not improbable since "submarines" and "submerged speeds" resemble each other in many ways. They both start with "s"; their first

syllables have central vowels; their last syllables have high front vowels% and so forth. If NP4 is to be applied, however, the RELN factor will resolve the doubt in favor of "submarine", and there will be no need to test in depth how well "submerged speed" maps onto the acoustic data or fits the semantic and discourse constraints.

The UNIT factor of NP11 guides the choice between "a" and "the", where acoustic evidence for a choice is typically lacking. Semantically, "a" resembles "one" in its ability to combine with numbers and units; e.g., "one ton", "a ton", "one hundred", "a hundred". If the instance of the NOM is "ton", "foot", "knot", or some other singular expression with the value UNIT for its CMU attribute, then "a" is judged to be more likely than "the". On the other hand, if the NOM is "fuel" or "submarines", the article cannot be "a". The CMU attribute for "a" is (COUNT UNIT), which does not intersect with the value (MASS) of the CMU attribute for "fuel"; the NBR attribute is (SG), which does not intersect with the value (PL) for "submarines". The factors referencing these attributes rule out application when the intersection is NIL. These are typical syntactic agreement tests.

As longer phrases are built up, the various attributes interact in other ways. For instance, the syntactic properties of relational expressions depend on which aspect of the relation is present in an accompanying prepositional phrase. Prepositional phrases have the attributes of their NP objects. When a prepositional phrase modifies a noun with the RELN

attribute, the CMU attribute for the resultant phrase is defined by taking the union of the values for the two nominal constituents. As a result, phrases like "surface displacement of the Lafayette" have the value (COUNT) and those like "surface displacment of seven thousand tons" have the value (COUNT UNIT). The difference in values marks the fact that the two examples do not fit with equal ease in all syntactic environments. It is referenced in the UNIT and RELN factors in (3) above, to influence the choice between the two articles, which are seldom distinguished clearly by sound. The rule is tuned to prefer "the" in the absence of the UNIT value; as in "the surface displacement of the Lafayette", and "a" when it is present, as in "a surface displacement of seven thousand tons". "A surface displacment of the Lafayette", which implies the possibility of having more than one surface displacement, is ruled out completely.

NPs also have a MOOD attribute, derived from their initial constituents. It is either declarative (DEC) as in "this submarine", or WH-interrogative (WH) as in "which submarine". The WH value is propagated to the larger phrases in which NPs are constituents. Sentences (S) and utterances (U) take the value for their MOOD attribute from an initial NP. Our current vocabulary does not include verbs like "know" and "tell", which can embed WH questions like "Do you know what the surface displacement is?" For the time being, we assume that noninitial noun phrases are not likely to have the value WH for MOOD. Echo

questions, e.g., "You said what?" are not ruled out, but have lower scores.

The convergence of many attributes at the higher levels of phrase composition makes possible many discriminatory judgments. Some of them are shown in (4).

(4)
```
RULE,DEF S3     S = NP;NP1 AUXB NP;NP2;

      ATTRIBUTES
       MOOD,FOCUS,CMU,RELN  FROM NP1,
       AFFNEG FROM AUXB;

      FACTORS
       NBRAGR1 = IF  CMU EQUAL "(UNIT) THEN
             [IF NBR(AUXB)EQUAL "(SG)THEN OK ELSE OUT]ELSE
              IF GINTERSECT(NBR(NP1),NBR(AUXB))THEN OK ELSE OUT,
       NBRAGR2 = IF  CMU(NP2) EQUAL "(UNIT)  THEN OK ELSE
           IF GINTERSECT(NBR(NP2),NBR(AUXB))THEN OK ELSE OUT,
       FOCUS = IF FOCUS(NP1) EQ "INDEF AND FOCUS(NP2) EQ "DEF
           THEN POOR ELSE OK,
       GCASE1 = IF GCASE(NP1) EQUAL "(ACC) THEN OUT ELSE OK,
       GCASE2 = IF GCASE(NP2) EQUAL "(ACC) THEN OUT ELSE OK,
       MOOD1 = IF MOOD EQUAL "(WH) THEN GOOD ELSE OK,
       MOOD2 = IF MOOD EQUAL "(WH) AND MOOD(NP2) EQUAL "(WH)
            THEN POOR ELSE OK,
       AFFNEG = IF MOOD EQUAL "(WH) AND AFFNEG EQ "NEG THEN
          BAD ELSE OK,
       RELN = IF RELN EQ "T AND CMU(NP2) EQUAL "(UNIT)
                THEN VERYGOOD ELSE OK,
       PERSAGR = IF GINTERSECT(PERS(NP1),PERS(AUXB))
                   THEN OK ELSE OUT;

      EXAMPLES
       THE LAFAYETTE IS A SUBMARINE (OK)
       THE LAFAYETTE IS SUBMARINES, WHAT IS THEM (OUT)
       A LAFAYETTE IS THE SUBMARINE (POOR)
       THEM ARE SUBMARINES, IT AM A SHIP (OUT)
       WHAT IS IT, WHAT IS THE LENGTH (GOOD)
       HOW MANY ARE WHAT (POOR)
       WHAT ISN'T THE SURFACE DISPLACEMENT (BAD)
       THE SURFACE DISPLACEMENT IS 7000 TONS (VERYGOOD);
```

The PERSAGR (person-agreement) factor tests for agreement between the so-called pronouns and the auxiliary constituent

The two grammatical case factors, GCASE1 and GCASE2, require that the grammatical cases of the two NPs are not accusative. These traditional syntactic agreement tests block application of the composition rule to putative expressions like "it are" and "they is", "Them is" is doubly blocked.

Some of the remaining factor statements in (4) are less traditional. One of these is the AFFNEG factor, which references both the MOOD and AFFNEG attributes and reduces the score greatly if the instance is purportedly a negative WH question like "what isn't the surface displacement?" Genuine requests for negative information occur in highly circumscribed situations. The rhetorical question is not a genuine request for information (e.g., "Who wouldn't like to be rich and famous!"). "Who isn't here?" is reasonable only if there is an established and limited list of people who are expected to be present, as in a classroom. "What isn't your name?" and "Where don't you live?" are patently absurd.

The constraint on negative WH questions is essentially due to pragmatic forces as well as semantic ones. Similar forces are at work in observed tendencies for the first NP in the composition defined by S3 to be indefinite in focus only when the second one is also. Stated oversimply, in coherent discourse, the things already talked about--the "old" information--tends to come first. What is predicated about it--the "new" information--tends to follow. Old information is information that has already been talked about and established in the

discourse, so that it is likely to be encoded in definite noun phrases. These are likely to be in subject position, so that the sentence they introduce is consistent with preceding sentences. New information tends to be introduced in indefinite noun phrases. The next mention of the "same thing" will then be old information, eligible for definite focus. Consequently, "A Lafayette is that submarine" seems peculiar, relative to "That submarine is a Lafayette". "A Lafayette is it" is still more peculiar. These discourse-based probabilistic tendencies are expressed in the FOCUS factor of S3.

The CMU attribute, as previously noted, is not purely syntactic. On the other hand, matters like number agreement have always been central to syntax. It is particularly interesting, therefore, that the number agreement constraints for S3 cannot be properly stated without appealing to CMU. To state number agreement constraints, Ns denoting units must be marked separately. Sentences like "These are a submarine", "These is a torpedo tube", "These is missile launchers", and "This are subs" are clearly ungrammatical, and the ungrammaticality is usually attributed to the fact that one of the constituents differs in grammatical number from the other two. However, "The surface displacement is seven thousand tons" is wholly grammatical even though two of the constituents are singular and the third is plural. Such use of semantic attributes in syntactic factors points to the conclusion that the integration of information from different sources of knowledge is well motivated on both

linguistic and heuristic grounds.

Because of the high frequency of WH questions in the protocols from which the vocabulary and phrase types were selected, the PG is now tuned to expect them. A sentence defined by S3 receives a higher score from the MOOD1 factor if its MOOD is WH. This tuning can easily be changed without altering the syntax or semantics of the language. If the user both extracts data from the data base and enters data into it, with no predictable pattern of alternation, factors like MOOD1 can simply be removed. A more interesting alternative is to reset them dynamically in a discourse context where the computer sometimes asks questions for the user to answer. After each user question, the grammar could be tuned to expect a declarative utterance whose syntax and semantics were appropriate and relevant.

References

Deutsch, B. G. Establishing Context in Task-Oriented Dialogs. Presented at the Thirteenth Annual Meeting of the Association for Computational Linguistics, Boston, Massachusetts, 30 October - 1 November 1975.

Hendrix, Gary G. Semantic Processing for Speech Understanding. Presented at the Thirteenth Annual Meeting of the Association for Computational Linguistics, Boston, Massachusetts, 30 October - 1 November 1975.

Paxton, William H., and Robinson, Ann E. System Integration and Control in a Speech Understanding System. Presented at the Thirteenth Annual Meeting of the Association for Computational Linguistics, Boston, Massachusetts, 30 October - 1 November 1975.

Walker, Donald E., et al. Speech Understanding Research. Annual Report, Project 3804, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, June 1975.