The Computer Generation of Speech with Discoursally and Semantically Motivated Intonation

Robin P. Fawcett

Computational Linguistics Unit
University of Wales College of Cardiff
Cardiff CF1 3EU
UK

Abstract

The paper shows how it is possible, in the framework of a systemic functional grammar (SFG) approach to the semantics of natural language, to generate an output with intonation that is motivated semantically and discoursally. Most of the work reported has already been successfully implemented in GENESYS (the very large generator of the COMMUNAL Project; see Appendix 1). A major feature is that it does **not** first generate a syntax tree and words, and then impose intonational contours on them (as is a common approach in modelling intonation); rather, it generates the various intonational features directly, as it is generating richly labelled structures (as are typical in SFG), and the associated items. The claim is not that the model proposed here solves all the problems of generating intonation, but that it points a way forward that makes natural links with semantics and discourse. A secondary purpose of this paper is to demonstrate, for one of many possible areas of NLG that could have been chosen, that there is still much important work to be done in 'sentence generation'. I do this in order to refute the suggestion, occasionally heard at recent conferences, that the major work in 'sentence generation' has already been done, and that the main (only?) area of significance in NLG is in higher level planning. In my experience the two are interdependent, and we should expect significant developments at every level in the years to come.

1. Purpose and Scope

The aspect of Natural Language Generation (NLG) to be described here is the generation of spoken text that has **intonation,** where that intonation is motivated both **semantically** (i.e. in terms of the semantics - in a broad sense of the term to be clarified soon - of sentences) and **discoursally** (i.e. in terms of what the discourse planning component specifies).[1] Any specification of intonation requires, of course, to be integrated with an adapted version of a **speech synthesizer** (e.g. one that draws on one of the currently available systems that attempts - inevitably with Appendix 1 for a brief overview of the project).

The approach is very different from that in MITalk (Allen 1986), which is essentially a text-to-speech system. So far as I am aware, the only generative model prior to ours that attempts to generate intonation that is motivated semantically and discoursally is the impressive work of Isard, Houghton, Pearson and their colleagues at Sussex (Houghton and Isard 1987) and Houghton and Pearson 1988). Its limitation is the very small size of its syntax, lexis, semantics and working domain. We see our work in the COMMUNAL Project as being to build on their important achievement. (But see Appendix 2 for what we do **not** attempt.)

2. The Relevant Components of the Communal Model

The major components of the overall model that will be referred to below are as follows. We assume an **interactive** system, rather than one that is merely **monologue.** We shall ignore here the components related to parsing, interpretation and inputting to the belief system and planner. The components relevant to generation are:

1. The **belief system,** which includes general and specific beliefs about ('knowledge of') situations and things in some domain; specific beliefs about the content of the preceding discourse, about various aspects of the current social situation, about the addressee and his beliefs of all types, his attitudes, his goals and plans.

2. A **planner,** which makes general plans, drawing on knowledge of ...

3. **genres** (scripts, schemas, etc), introducing where appropriate sub-units such as **transactions** (see below) and more detailed plans, using ...

4. the **local discourse grammar,** which is modelled as a 'systemic flowchart' (i.e. a flowchart containing many small system networks at the choice points, and which generates **exchanges** and their structure),

5. the **lexicogrammar,** i.e. the **sentence generator,** consisting of:

[1] As will be clear from what follows, the model presented here owes a great deal to two people in particular: Michael Halliday and Paul Tench. Through them, I am well aware, there is a debt to many others, too numerous to mention, who have worked in the difficult field of intonation in English. I am grateful too for early encouragement in this area from Gillian Brown (whose work is drawn on also by Paul Tench), and for the regular, ongoing stimulus of many good explorations of ideas in this and other areas with my colleague Gordon Tucker. But none of these should be blamed for the inevitable crudities, infelicities and no doubt errors in the model described here; these are mine.

a. the **system networks of semantic features** for a wide variety of types of meaning related to **situations** and realized in the clause, including **theme** and **information structure** as well as **transitivity, mood, negativity, modality, affective meaning** and **logical relationships**, and equivalent system networks for **things** and **qualities**, and

b. the **realization rules** which turn the **selection expressions** of features that are the output from passes through the system networks into **syntactic structures** with **items** (grammatical and lexical) and markers of **punctuation** or **intonation** as their terminal nodes.

3. Modelling Intonation

3.1. An Overview of the Generation of Intonation

Let us imagine that Ivy (the ’person’ of whose mind GENESYS models a part) is about to generate a sentence. Let us suppose that she is being consulted by the Personnel Officer of a large institution, who draws regularly on her specialist knowledge and advice, and that he has just asked her "Where does Peter Piper live?". (We shall come later to how intonation is represented.) Like most human users of language, Ivy makes reasonable assumptions about (loosely, she ’knows’) where she is in any current **transaction** (e.g. at the start, in the middle or at the end), and where she is in the current **exchange**. This affects the pitch level of what she says. She needs to choose a **tone** (the change in pitch marked by a stepping or a slide on the **tonic** syllable) which will express the **MOOD** of the final matrix clause of her sentence. (’Matrix’ here means ’at the top layer of structure’.) She needs to locate that tone on an item which will be thereby marked as **new information**. She needs to decide if it is to be presented simply as ’new’, or as ’contrastively new’ (in the terms used here). And she needs to decide on the **information status** of any chunks of information that are to be presented as separate from the main **information unit** of the clause. (The information that guides these choices comes from various aspects of the higher belief system, which there is unfortunately no space to discuss here.)

As we shall see, these various components of the semantic level of intonation account, in a different way from the usual approach in British intonation studies, for Halliday's well-known triad of TONE, TONALITY and TONICITY. While it is perfectly possible to talk about the contrasts in intonational form to which these three refer as ’systems’, I suggest that it is more insightful to take, as the level of contrasts to be modelled in system networks, the **meanings** that lie behind (or, in the SFG metaphor, above them). These **semantic** features are then **realized** in the purely intonational contrasts of TONE, TONICITY and TONALITY.

The accounts of the various aspects of intonation in what follows will inevitably be introductory, and may to the specialist appear simplistic. A somewhat fuller treatment is given in Tench and Fawcett 1988, and a very full treatment is given in Tench 1987, which includes summaries of relevant work by other intonation specialists.

(I omit here, for reasons of space, a specification of how one might model the way in which the position in a **transaction** and an **exchange** affects intonation.)

Let us assume, then, that Ivy is preparing a response to the Personnel Officer's question, using the information that, while Mr Peter Piper's address is currently 11 Romilly Crescent, Canton, Cardiff, he is moving from there after one month. In discourse planning terms, she chooses that her move will be a ’support’ for a ’solicit information’, and that the act at the head of the move is a ’give new content’ (see Fawcett, van der Mije and van Wissen 1988). As we shall see, these choices **pre-select** in the MOOD network of the **lexicogrammar** the features [information] and [giver]. But first there is a more basic system to consider.

3.2. The MODE System

The initial rule of the semantics of the lexicogrammar to be considered here is:

situation ->
’MODE’ & ’TENOR’ & ’CONGRUENCE_SIT’.

Thus means that, for any ’situation’ (roughly = ’proposition’) that you are generating, you must make choices in all three of the systems named. (Notice, then, that ’parallelism’ lies at the heart of the grammar.) Here we shall be concerned only with the MODE system. (It is from CONGRUENCE_SIT that the main part of the network is entered, to generate configurations of participant roles, such as Agent and Affected, and choices in MOOD, such as ’information seeker’, and very many others.) The MODE system is very simple:

’MODE’ -> 70% spoken / 30% written.

This means: ’In the MODE system, you must choose between generating a spoken output (for which under random generation there is a 70% probability) and generating a written output (which carries a 30% probability). Clearly, since

Ivy is in a spoken interaction, she will be strongly disposed to select [spoken] - but in principle she need not. We shall not discuss here the interesting reasons for and against introducing this system to the lexico-grammar itself, except to point to two significant advantages that it brings. Nor, unfortunately, is there space to discuss the roles of the probabilities and the ways in which they are assigned (sometimes simply a guess at the overall pattern for central types of text; sometimes based on textual studies).

(The next few lines presuppose some familiarity with systemic grammars; for those without this knowledge it may be advisable to re-read this section after seeing the working of the examples.) What is the role of this system? First, it enables the grammar-builder to refer, at any point on this initial pass through the **system network**, to whichever feature in this system has been chosen as an **entry condition** to a later system. In other words, where there is greater richness of choice in meaning in the spoken mode (as is typically the case with meanings realized in intonation, as against those realized in punctuation), we can ensure that those systems are only entered when the feature [spoken] has been chosen. We shall shortly see the great value of this. Second, the 'MODE' system enables us to refer, at any point in a **realization rule**, on this **or any subsequent pass** through the network, to this feature as a **conditional feature** for the realization of some other feature. In other words, we can ensure that if [spoken] has been chosen the realization will take the form of intonation, and if [written] has been chosen it is expressed in punctuation. Both of these facilities contribute greatly to the elegant operation of the lexicogrammar as a whole, both in meanings realized in intonation and in many other ways.

3.3. The Sentence Generator: MOOD

The unmarked choice in the 'CONGRUENCE_SIT' system is, unsurprisingly, [congruent_sit]. This is the choice that opens up the whole array of meanings associated with realization in a clause, and many parallel systems follow. Among these is the MOOD network. This is a fairly large and complex network of meanings, and these are realized partly in **syntax**, partly in **items** (such as "please"), and partly in **tone** (= variation in pitch). The network is too large and complex to present here, but we shall trace a route through it that shows why it is central to an understanding of intonation. The first options in the current GENESYS network are shown below:

```
congruent_sit ->
  'TRANSITIVITY' & 'MOOD (& OTHERS).

'MOOD(A)' -> 90% information / 10% directive.

information -> 'MOOD(B)' &
'TIME_REFERENCE_POINT' (& OTHERS).

'MOOD(B)' ->
  70% giver (1.2) / 30% seeker (16).
```

The second line reads: ' In the MOOD(A) system you must choose between the feature [information], which overall has a 90% probability of being selected, and [directive], which has only a 10% probability. As so often, the choice of a single feature leads to further parallel systems, one of which continues the MOOD network itself. The last line in the above rules exemplifies the use of numbers in brackets after the features; it is the number of the **realization rule** for the feature concerned. What will this look like?

Here is a slightly simplified version of the realization rule for [giver]:

```
1.2 : giver :
  if fills 'Z' and (simplex_sit or
final_co_ordinated_situation)
  then (if on_first_pass written then 'E' < ".",
        if on_first_pass spoken then 'E' < "|").
```

The effect of this rule is on the 'Ender' (i.e. 'E', the last element in the structure of the clause). If [written] is chosen in the 'MODE' system it is expounded by a full stop (Br. E. for 'period'), but if the choice is [spoken] it is expounded by a final intonation unit boundary, i.e. |. However, says the rule, neither realization will occur unless the clause (1) directly fills the element 'Sentence' (represented by 'Z' as an approximation to sigma) and (2) is 'simplex', i.e. is not co-ordinated with one or more other clauses or, if it is, is the final clause.

This may seem a surprisingly complex rule to those in NLP used to working with mini-grammars. But this is typical of the working level of complexity in a natural language, and those who are used to working with the problems of building broad coverage grammars will appreciate that this is not a particularly complex rule. In the case of our example the effect is to give to Ivy's output a final intonation unit boundary.

We come next to an example of the value of being able to use of the feature [spoken] as an entry condition to a system. This is necessary because the MOOD network also builds in variables in 'key' (in the sense of Halliday 1970), i.e. finer choices within the MOOD options. These correspond to what Tench treats separately as variations in **attitude**. While accepting the view that these more delicate choices can be seen as serving a separate function from the function of the basic tone, the fact is that in any systemic computational implementation the way in which they enter the choice system is simply as more delicate choices that are directly dependent on the broad choice of meaning realized in the broad tone. The range of such delicate variations appears to be potentially different for the various meanings (see further below). In the systems given below, note the high probability of choosing [assertive] followed by [neutral].

giver & spoken ->
  70% assertive / 15% deferring (1.21) /
  15% with_reservation (1.22).

assertive -> 2% very_strong (1.23) / 8% strong
  (1.24) / 60% neutral (1.25) / 30% mild (1.26).

In an intermediate level model (such as Prototype Generator 2 (PG2), which is the most advanced version of GENESYS currently implemented) we need only relatively simple rules such as the following:

1.21 : deferring:
  if fills 'Z' and on_first_pass spoken and
  (simplex_sit or final_co_ordinated_situation)
  then '2-' by 'NT'.
    ('NT' = 'Nuclear Tonic'; see 3.4.3.)

1.22 : with_reservation :
  if fills 'Z' and on_first_pass spoken and
(simplex_sit or final_co_ordinated_situation)
then '12' by 'NT'.

And so on, for [very_strong] (realized by '21'), [strong] (realized by '1+'), [neutral] (realized by '1') and [mild] (realized by '1-').

Here we are using a numerical notation for tones that goes back to an earlier tradition even than Halliday's description (1967, 1970), though it has much in common with Halliday's. (Readers from the American tradition used to an iconic representation may have some adjustments to make in interpreting the notation. But there should be no fundamental difficulty; Halliday's description has been widely used (and indeed tested) on American and Canadian English.)

I give next a brief summary of the differences between the scheme for tones used here and Halliday's well-known scheme (1967, 1970). Tench's (and so my) numbers '1' and '2' correspond to Halliday's usage, as do the use of '+' and '-'. But Halliday's 'Tone 3' is seen as a variant of our Tone 2; Halliday's Tone 4 (a fall-rise) is represented by '12'; and his 'Tone 5 (a rise-fall) is shown as '21'. Tench's general descriptions of the tones in words (1987) imply four pitch levels, and I therefore use the following labels for the model implemented: base, low, mid and high. The four levels in turn provide a framework for describing three types of pitch change. It will be helpful for what follows to set them out as three 'scales'; these descriptions of the tones are in effect source material for writing realization rules. (I have given these scales informal semantic labels; these are not intended to correspond directly to the features in the MOOD network encountered so far, but to evoke features from various parts of the network, including the many options dependent on [directive]). Finally, let me remind you that we are not at this point trying to account for all tones, but only for those that carry the MOOD of a matrix simplex or final clause. This clear separation of the ways in which tones are generated is a key feature of the present proposals. We shall come shortly to some of the ways of generating appropriate tones for some of the other positions in which tones occur.

The 'assertive' scale (Tones 21 and 1):
  Tone 21: rise-fall (rise to high plus fall to base)
  Tone 1+: high-fall (fall from high to base)
  Tone 1: mid-fall (fall from mid to base)
  Tone 1-: low-fall (fall from low to base)
Also (see below):
  Tone 21-: low rise-fall (lower version of Tone 21)*

The 'deferring' scale (Tone 2):
  Tone 2+: high-rise (rise from base to high)
  Tone 2: mid-rise (rise from base to mid)
  Tone 2-: low-rise (rise from base to low)

The 'implication' scale (Tone 12):
  Tone 12: fall-rise
Also (see below):
  Tone 12-: low fall-rise (lower version of 12)*

* Tench suggests that these are variants of Tone 21 and 12 that additionally signal 'emotional involvement'.

As will be clear, Tench and I propose a modification to Halliday's basic set of contrasts in TONE, such that Halliday's Tone 5 is seen as an extreme form of Tone 1. This fits naturally with the semantics of these tones. In a somewhat similar way, Tench treats Halliday's Tone 3 (a low rise) as a variant of Halliday's Tone 2, under the rubric of 'deference to the listener', and we adopt this too in COMMUNAL. But note that, while that kind of semantic description holds good for Tone 2s (Halliday's Tone 3s) in the sentence-final position, I shall suggest other means of generating them in non-final positions. In the present system there are no 'double tone groups', such as Halliday's Tone 13 (i.e. a Tone 1 to realize the main MOOD meaning, followed by a Tone 3 (here 2) for 'supplementary information'.) Such final Tone 2s will be generated in a similar way to that to be illustrated in section 3.5 below for initial Tone 3s (and, as we shall see, 12s). Finally, note that I include here one option that Tench includes under 'status of information'. This is his 'implication', realized in Tone 12, i.e. a fall-rise. This is Halliday's Tone 4, which he characterizes as (among other descriptions) 'with reservation'. This tone occurs both as a carrier of MOOD and otherwise; it is with the former that we are concerned here. It seems plausible to treat it as a variant that can be chosen as an alternative to the basic falling and rising tones recognized by both Halliday and Tench, and I have therefore incorporated it in the overall MOOD network.

3.4. Focus of Information

3.4.1. The Line of Approach to the Problem

I shall present here a somewhat novel approach to the relationship between the two sets of phenomena described by both Halliday and Tench as **TONALITY** and **TONICITY**. TONALITY is typically thought of as 'cutting up a string of words into intonation units' (Tench's term; Halliday's is 'tone groups'), with each **intonation unit** realizing one **information unit**. The problem, when one is approaching the question from the angle of generation, is that **there is no string of words to cut up** - not, that is, until the sentence has been generated. We therefore need to look for a **semantic** approach to the problem. My proposal is that it is helpful to start not with TONALITY but TONICITY.

TONICITY is the placing of the **tonic** on a syllable. The item so marked is shown to be being presented as **new information** - and this is a semantic concept. ('New' information is information presented as 'not recoverable.') But a further problem arises, in that linguists recognize both 'marked' and 'unmarked' tonicity.

3.4.2. Generating Marked Tonicity

Marked tonicity occurs when the item containing the tonic syllable is presented by the speaker as 'contrastively new'. **Unmarked tonicity** occurs when there is no marked tonicity (which is by far the most usual case); we shall return to this shortly. Marked tonicity is handled in GENESYS in the following way. In principle, any pathway through the system network that results in the generation of a **formal item** will lead to a system of the following form (where 'x' is the current terminal feature):

x -> not_contrastively_new / contrastively_new.

The realization of [contrastively_new] is that a **contrastive tonic** is conflated with the **element of structure** that the item expounds. The simple version implemented in PG2 is as follows:

'INFORMATION_FOCUS' ->
99% no_element_marked_as_contrastively_new /
1% element_marked_as_contrastively_new.

element_marked_as_contrastively_new ->
 50% contrastive_newness_on_polarity (18.1) /
 50% contrastive_newness_on_process (18.2) /
 0% other.

Realization rule 18.1 states the complex set of conditions for conflating a **contrastive tonic** ('CT') with the appropriate element; for the POLARITY system ('positive' vs. 'negative') this is typically the Operator (which may have to be supplied by a 'do-support' rule) but it may be any one of several others, depending on whether or not the clause is moodless and, if not, whether a directive, and if not, what auxiliaries are realized, etc. The rule for presenting the 'process' (realized in the Main verb) as 'contrastively_new' is however extremely simple:

18.2 : contrastive_newness_on_process :
 'CT' by 'M'.

In the case of our example, the choice is not to present any element as contrastively new.

3.4.3. Generating Unmarked Tonicity

How, then, should we generate **unmarked tonicity**? The answer is simple: as the default - i.e. when there is no contrastive tonic. In other words, I want to suggest that unmarked tonicity is a formal phenomenon of intonation that does **not** express an active choice in meaning. The relevant facts are well known, i.e., roughly, that what we here term a **nuclear tonic** ('NT') falls on the last lexical item in the information unit. The question is: 'How can we define the intonation unit, in semantic terms?' The only contender as a semantic unit, in the GENESYS framework, is the **situation**, i.e. the semantic unit typically realized in the clause. The actual decision as to which item the unmarked tonic shall be assigned to gets made relatively late in the generation process. In GENESYS we simply have a list of the few dozen items generated by the lexicogrammar that cannot carry the unmarked tonic: roughly, the 'grammatical items' of English. Essentially, then, this default rule will insert **one**, and only one, nuclear tonic in each sentence. This will hold even when there are two or more co-ordinated clauses in that sentence, and/or one or more embedded clauses.

3.5. Status of Information

3.5.1. The Importance of this Category

This is a concept not distinguished as a separate phenomenon in Halliday's treatment of intonation, but which Tench does treat separately. This clear separation of two semantically distinct phenomena was a significant help in developing the generative model proposed here. However the concept of 'status of information' is quite highly generalised, in the sense that it is not manifested in just one part of the overall network (as for example MOOD is). Specifically, we find this option at many of the points **where a unit is generated that is not the final matrix clause in the sentence.** Many of these (though by no means all) have already been implemented in GENESYS, and the following are a representative sample.

3.5.2. The Co-ordination of Situations and Things

One major source of multiple intonation units is **co-ordination.** Thus, when GENESYS generates co-ordinated **clauses** (realizing co-ordinated **situations**) such as "Either Ivy loves Ike, or she loves Fred, or she doesn't love anybody.", she first recognizes at an abstract level that separate information units are being assigned and then inserts, depending on whether the output is to be spoken or written, either (1) commas or (2) intonation unit boundaries and an appropriate tone such as Tone 2. We shall not reproduce here the surprisingly large system network and realization rules for this area of the grammar, which merit a paper to themselves. All that needs to be said is that to develop a model of clause co-ordination that incorporates most of the phenomena of naturally occurring texts is a major task, and that it took several months of work to build our current system. In terms of the above example, it generates, if [spoken] has been selected:

```
| either Ivy loves Ike/NT/2 |
  or she loves Fred/NT/2 |
  or she doesn't love anybody/NT/1 |
```

While the patterns of the networks and their realizations are different for the co-ordination of nominal groups, they are handled in a similar way. The system accommodates the perhaps surprising fact that, in the case of nominal groups, there is typically one more intonation unit than there would be commas. As in the MOOD network, there is a greater number of delicate choices realized in intonation than there is in punctuation. So the feature [spoken] is again used as an entry condition to the system in the 'CO_ORDINATION_SIT' network, to ensure that the system is not entered unless [spoken] has been chosen. Here the speaker chooses in the system of [unmarked_co_ordination_spoken] vs. [co_ordination_with_reservation]. The first is realised by a Tone 2 base-to-mid rise; Halliday's Tone 3), and the second by a Tone 12 (fall-rise; Halliday's Tone 4).

3.5.3. Thematized Circumstances: Situations, Things, Qualities

Another major source of additional intonation units is the **thematization of time and circumstance.** These meanings are realized in Adjuncts of various types. They may occur in various places in the clause, and here we shall consider just those that appear at the beginning of a clause. So far GENESYS includes eleven types, each of which may be realized by either a **clause** or a **group** (three different classes of groups being recognized: **nominal, prepositional** and **quantity-quality groups**). Note, then, that we have now identified a second major source of what has been termed 'clause combining'. A similar approach is needed for 'clause final' clauses, i.e. clauses that fill any of the eleven types of Adjunct built into GENESYS so far, and that come late in the clause. (This is a different approach to clause-combining from that in Halliday 1985 and so from that in the Nigel grammar at ISI; here such clauses are simply treated as embedded - so far with gains in generalizations rather than losses.)

Let us take as an example the concept of **time position,** which is one of five types of 'circumstance of time' recognized in GENESYS - the others being **repetition, duration, periodic frequency,** and **usuality.** While GENESYS will happily generate clauses such as "until he leaves the company" to specify a time position, in the case of our example Ivy has chosen the simpler structure of the prepositional group, i.e. "until next month". The first system to consider is:

```
'TIME_POSITION_THEMATIZATION' ->
   99% unthematized_time_position (20.2) /
   1% thematized_time_position (20.3).
```

Because the answer modifies the presuppositions that the Personnel Officer brought to his question (i.e. that Peter Piper had a fixed address), Ivy decides to **thematize** the part of her reply that expresses this, i.e. her specification of the 'time position'. This is realized by placing the 'time position Adjunct' at an early place in the clause. (Note that this is not a 'movement rule'; there are no such rules in this generator, and no element is located until it can be located in its correct place.) The next two systems are:

```
thematized_time_position ->
  80% time_position_as_separate_
    information_unit /
  20% time_position_as_part_of_main_
    information_unit.

spoken & time_position_as_separate_
  information_unit ->
  20% highlighted_thematized_time_position /
  80% neutral_thematized_time_position.
```

The first of the two systems applies whether or not the MODE is spoken or written (the written realization being a comma). But the writing system cannot make the distinction offered in the second, so that here again the feature [spoken] from the original MODE system is used as an entry condition. In our example Ivy chooses to present the specification of the time position ("until last week") as a separate information unit, and furthermore to highlight it (by using a Tone 12 (a fall-rise).

But you may have noticed that these features have no realization rules. How, then, do these choices get realized? The answer is that these features act as **conditional features** on the realization rules for the units that are generated, after re-entry to the overall network, **on a subsequent pass** though it. The reason for including the system at the rank of situation is that in this way we can capture the generalisation that these options are relevant **whatever the unit** - a clause or some kind of group - that fills the Adjunct.

In our case the sub-network that we find ourselves in on re-entry is the network for 'MINIMAL_RELATIONSHIP_PLUS_THING', i.e. the network from which **prepositional groups** are generated. Here we enter the following system (where the suffix 'mrpt' echoes the name of the overall system):

```
  location_mrpt ->
    place_mrpt (90.001) / time_mrpt (90.002).
```

Here [time_mrpt] will be **pre-selected** by the choice at the higher rank. The part of its realization rule concerned with intonation may appear, once again, somewhat complex, but once again it seems to correspond to the relative (but always limited) complexity of the facts of how English works:

```
90.002 : time_mrpt :
   if   (on_previous_pass
time_position_as_separate_information_unit
and on_first_pass spoken)
  then (if current_unit pgp then e < "|"),
       if  on_previous_pass
highlighted_thematized_time_position
       then '12' by 'T',
       if  on_previous_pass
neutral_thematized_time_position
       then '2' by 'T'.
```

As you will see, these rules insert appropriate intonation boundaries and tones. The tonic ('T') is already waiting in the **starting structure** of the prepositional group, so that the rule simply conflates the actual tone with it. Let us assume that Ivy, in order to highlight still further the thematization of the words " month", selects the **highlighting** rather than the neutral option. (The nominal group "next month" is generated by a further re-entry.) Finally, the system supplies the initial intonation unit boundary for any unit without one. If we assume that the rest of items generated (in components not considered in this paper) are "he will be living at eleven Romilly Crescent, Canton" the full output for our example is:

| until next month/T/12 | he will be living at
  eleven Romilly Crescent/T/2 | Canton/NT/1 |

3.5.4. Other Sources of Intonation

Other sources of intonation occur in specialist mini-grammars such as those for dates and addresses. These can be quite complex, and may insert several tonics, each with an appropriate tone. Our worked example illustrates one such case: note the Tone 2 on "Crescent". Yet other types will be included in the next version of GENESYS, including (1) Adjuncts (which may be filled by clauses or groups) that are placed **after** the nuclear tonic of a clause and which carry 'supplementary information', and (2) 'non-restrictive relative clauses' (i.e. ones that carry, once again, 'supplementary information').

3.6. Summary of the lexicogrammatical generation of intonation

We have now completed a fairly full specification of the major aspects of intonation included at the present stage of the development of the GENESYS model.

To summarize: GENESYS offers the choice, on entering the first system that results in the generation of a sentence, between [written] and [spoken]. The importance of this apparently trivial system is that the choice made in it determines whether or not one can go on to enter quite a number of more 'delicate' systems whose choices are realized in intonation. Its features also act as **conditions** on the realization of features chosen in the same network, or in one entered on a subsequent pass. The result is that the realization at the level of **form** will be in terms of **either** intonation **or** punctuation. We have seen how choices in MOOD, in INFORMATION_FOCUS and in various types of 'status of information' contribute together to the specification of intonation, and we have seen some of the details of how this can be implemented. The result is an integrated model that avoids the psychologically implausible approach whereby one **first** generates a syntax tree and a string of words at its leaves, and **then** 'adds on' the intonation. Instead, it treats intonation as one of three modes of realization (the other two being syntax and items), generating the various aspects of intonation at appropriate points in the generation of syntax and items.

It may be helpful to conclude by specifying explicitly the final stages of this process. **First** the generator looks for a **contrastive tonic** ('CT') with which to conflate the tone, and **then**, if there isn't one, it provides as a default a **nuclear tonic** ('NT') for the final matrix clause, i.e. the intonational element of structure with which the tone realizing the meaning of MOOD is conflated. The other intonation units specified by various types of **information status** are fitted around this central framework, receiving tones appropriate to their status. Where they are clauses these tones will be conflated with a nuclear tonic (unless, of course, there is a contrastive tonic), and where they are groups the tones will be conflated with a simple tonic. A nuclear tonic is thus one that is potentially capable of receiving the type of tone that realizes a MOOD option. It should be made clear that, in every case of the location of a nuclear tonic or a simple tonic, the element with which it is conflated must be one that is not expounded by an item from the list of inherently weak items. (Any such item may of course still receive a tonic by being contrastively stressed, as in | he has/CT/1+ eaten it |.)

4. Conclusions

4.1. Overall Summary

The COMMUNAL project began with a hope that it would be possible to take the insights from a Hallidayan-Tenchian view of intonation, and to develop a computational adaptation and implementation of them. A promising overall approach to the problem has indeed been developed; much of the resulting model has been worked out in considerable detail; and many large and significant portions have been implemented computationally. The framework has proved itself to be adaptable when modifications are indicated, and there is good reason to hope that aspects not as yet worked out explicitly will prove to be solvable in the framework of the present model. There is, therefore, the exciting prospect that, when our sister project gets under way and provides the necessary complementary components (no doubt with some requirements on us to adapt our outputs to their needs) we shall be in a position to offer a relatively full model of **speech with discoursally and semantically motivated intonation.** It will, moreover, be a principled model, and we hope that it will be capable of further extension and of fine-tuning. We feel that the use of SFG, and specifically of the type that separates clearly system networks and realization rules (as in GENESYS), gives us a facility that is sensitive to the need for both extension and fine-tuning. Above all, the centrality in the model of **choice** between **semantic features** makes it a natural formalism for relating the 'sentence grammar' to higher components in the overall model.

4.2. The General Prospect in NLG

Finally, let me turn to a more general point. It appears that, increasingly over the last few years, the focus of interest for many researchers in NLG has switched from what we might term **sentence generation** to higher level planning (which I term **discourse generation**). It is here, one sometimes hears it said, that 'all the really interesting work' is being done. Going implicitly with this claim is the assumption, which I have occasionally heard expressed quite explicitly, that the major problems of sentence generation have been solved.

But is this really so? While a lot of very impressive work has been done, and while some quite large generators have been built (e.g. as reported in McDonald 1985, Mann and Mathiessen 1985, Fawcett 1990), very many major problems remain unresolved. Specifically, many important aspects of 'sentence grammar' remain outside the scope of current generators. Where, for example, will we find a full description of a semantically and/or pragmatically motivated model of even such a well-known syntactic phenomenon as the relative clause? And what about comparative constructions (where even the linguistics literature is weak)? And there are many, many more areas of the **semantics and syntax of sentences** where our models are still far from adequate. There are also many **issues of model construction** regarding, for example, the optimal division of labour between components, the outlining of which deserves a separate paper (or book). And, even if we had models that covered all these and the many other areas competently, we have hardly begun the process of developing adequate methods for the **comparison and evaluation** of models. Thus there is still an enormous amount of challenging and fascinating work to do before we can say with any confidence that we have anything like adequate sentence generators. (A senior figure in German NLP circles suggested at COLING '88 that one can buy good sentence generators off the shelf. It depends how good 'good' is!)

In this paper I have illustrated two crucial points: (1) that there are indeed significant areas of language not yet adequately covered in current generators, and (less clearly because I have had to omit the relevant section for reasons of space) (2) that the development of an adequate model of these depends on the **concurrent** development of discourse and sentence generators.

Clearly, while there are in existence a number of fairly large sentence generators, we have in no way reached a situation where no further work needs to be done. I am aware, as the director of a project that seeks to provide rich coverage for as much of English as possible, that we have a great deal of work still to do, and that this holds for the sentence generator component as well as for the discourse planning systems. GENESYS already has 50% more systems than the NIGEL (in the long established Penman Project; see Appendix 1), but our rough estimate is that we need to make it at least as large again before we have anything approaching full grammatical coverage. And, of course, as everyone who has wrestled seriously with genuine natural language knows, many tricky problems will remain even then. Finding anything like the 'right' solution to many of these will require, I claim, models that have developed, in close interaction with each other, their discourse planning and their sentence generation components - and their belief representation, including beliefs about the addressee.

Acknowledgements

The research reported here was supported by grants from RSRE Malvern under contract no. ER1/9/4/2181/23, by the University Research Council of International Computers Ltd, and by Longman.

Appendix 1

COMMUNAL is a major research project that applies and develops **Systemic Functional Grammar** (SFG) in a very large, fully working computer program. The acronym COMMUNAL stands for **COnvivial Man-Machine Understanding through NAtural Language**. The principles underlying the project are set out in Fawcett 1988, and an illustration of a generation is presented in Tucker 1989. A fuller (but fairly informal) overall description, including some comparison with other projects, is given in Fawcett 1990. See also Fawcett (to appear). The project is planned to last 5 years, with around 6 researchers working on it. We finished the successful Phase 1 in 1989, and now (May 1990) are getting under way on Phase 2 The central component of the overall system is the **generator**, built at Cardiff. This is called GENESYS (because it GENErates SYStemically). Contributions from the University of Leeds in Phase 1 were to build (1) a derived **probabilistic parser**, called the RAP (for Realistic Annealing Parser, which develops earlier work at Leeds), and (2) the **interpreter** (called REVELATION, because it reveals the 'meaning' from the 'wording'). Each of these is a major development in its field. But because both build directly on the relevant aspects of GENESYS, we can characterise the **coverage** of the COMMUNAL system as a whole in terms of the size of GENESYS.

Here is a quotation and a few facts to give you a perspective on COMMUNAL at the end of Phase 1. McDonald, Vaughan and Pustejovsky (1987:179), in referring to the Penman project at the University of S. California, say: 'Nigel, Penman's grammar .... is the largest systemic grammar and possibly the largest machine grammar of any kind.' Although the COMMUNAL team developed GENESYS completely independently, starting from scratch with new system networks and handling realization in a rather different way, GENESYS already has many more systems than Nigel. (This is not a criticism of Nigel; the research team have been working on other components of Penman). A major theoretical difference between the two is that the networks in GENESYS are more explicitly oriented to semantics than in Nigel. We make the assumption that the system networks in the lexicogrammar **are** the semantic options. GENESYS has around **600 semantic systems realized in grammar** (syntax and morphology, and also intonation and punctuation (see below), while Nigel has about 400 grammatical systems. But GENESYS additionally does something that the builders of Nigel would have liked to do, but from which they have so far been prevented (by the requirement of a sponsor): it integrates **system networks for vocabulary** with the networks realized grammatically. GENESYS is still growing, so that in Phase 2 we estimate that it will more than double the number of systems realized in syntax and grammatical items. This should enable it to handle something approaching unrestricted syntax. COMMUNAL's first major achievement is therefore the size and scope of

GENESYS. The second must be seen in the wider framework of the model as a whole. It has been a long-standing goal of NLP to build a large scale system that uses the **same grammar** to either **generate** or **interpret** a sentence. (Many current systems use a different grammar for each process.) The second major achievement is to have performed this task with a very large grammar - a Systemic Functional Grammar, in this case. (This will be the subject of a separate paper in the future.) A third achievement (though one less relevant in the present context) has been the development of a **probabilistic parser** by the Leeds part of the COMMUNAL team.

Appendix 2

'Intonation' is a term susceptible to a wide range of interpretations. It may therefore be useful to list some major aspects of the complex task of generating natural intonation that will **not** be discussed here. The first four are not covered because they lie outside the current goals of the COMMUNAL project, while the last two are omitted because they will be implemented (we expect) by a sister project, support for which is currently being negotiated.

1. We shall not be concerned with the high level planning that will **tailor** the text to the needs of the addressee as affected by the channel (e.g. to build in greater redundancy, in the form of repetition of subject matter in planning what to express overtly, act by act). (For the general notion of tailoring, see Paris 1988.)

2. We shall not discuss variation in intonational characteristics of the sort that distinguish between speakers of different **dialects** (geographical, social class, age, etc).

3. The same goes for individual variation, i.e. intonational **idiolect**.

4. We shall ignore the code of **tone of voice** ('angry', 'conciliatory', 'delighted', etc). At the same time we recognize that it is an important semiotic system in its own right, and that in the longer run the way in which it is, as it were, superimposed on the intonation system itself must be modelled. We recognize too the problems of drawing a firm line between tone of voice and some of the quite delicate distinctions that we shall recognise in the MOOD system (c.f. Halliday's 1970 term 'key').

5. We shall ignore any aspect of **intonational variation that does not realize meaning**. For example, it may be that speakers introduce semantically unmotivated variation into the pretonic segment of an intonation unit, in order to avoid monotony (cf. House and Johnson 1987). (An alternative hypothesis, of course, might be that such variation is in fact semantically motivated, but that we have not yet discovered what aspects of meaning it correlates with and how best to refer to it; this is a characteristic of much **interpersonal** meaning.)

6. We shall not be concerned here with the **physical implementation** of the output, but simply (if only it were simple!) with providing a written text output marked appropriately for input to the system which will integrate it with the speech synthesis representation of the segmental phonology.

References

Allen, J. (ed.) *From Text to Speech: the MITalk System.* Cambridge: Cambridge University Press.

Benson, J.D., and Greaves, W.S., (eds.) 1985. *Systemic perspectives on discourse, Vol 1: Selected theoretical papers from the Ninth International Systemic Workshop.* Norwood, N.J., Ablex.

Brady, M., and Berwick, R.C. (eds), 1983. *Computational models of discourse.* Cambridge, Mass: MIT Press.

Fawcett, R.P., 1980. *Cognitive linguistics and social interaction: towards an integrated model of a systemic functional grammar and the other components of an interacting mind.* Heidelberg: Julius Groos and Exeter University.

Fawcett, R.P., 1988. 'Language generation as choice in social interaction'. In Zock and Sabah (eds) 1988b, 27-49.

Fawcett, R.P., 1990. 'The COMMUNAL Project: two years old and going well'. In *Network No. 13.*

Fawcett, R.P., (to appear). 'A systemic functional approach to selectional restrictions, roles and semantic preferences'. Accepted for *Machine Translation*.

Fawcett, R.P., van der Mije, A., and van Wissen, C., 1988. 'Towards a systemic flowchart model for local discourse structure', in Fawcett and Young 1988, pp. 116-43.

Fawcett, R.P., and Young, D.J., (eds.) 1988. *New developments in systemic linguistics, Vol 2: Theory and application.* London: Pinter.

Halliday, M.A.K., 1967. *Intonation and grammar in British English*. The Hague: Mouton.

Halliday, M.A.K., 1970. *A course in spoken English: intonation*. London: Oxford University Press.

Halliday, M.A.K., 1985. *An introduction to functional grammar*. London: Arnold.

Houghton, G. and Isard, S.D.,1987. 'Why to speak, what to say and how to say it: modelling language production in discourse'. In Morris 1987, pp. 112-30.

Houghton, G., and Pearson, M., 1988. 'The production of spoken dialogue'. In Zock and Sabah 1988a, pp. 112-30.

House, J. & Johnson, M. (1987) 'Enlivening the intonation in Text-to-Speech Synthesis: an 'Accent-Unit' Model', *Procs 11th ICPhS*, Tallinn.

Kempen, Gerard, (ed) 1987. *Natural language generation*. Dordrecht: Martinus Nijhoff.

Kobsa, A., and Wahlster, W. (eds.) *User Models in Dialogue Systems*. Berlin: Springer.

Mann, W.C., and Matthiessen, C.M.I.M., 1983/85. 'A demonstration of the Nigel text generation computer program'. In Mann and Matthiessen 1983 and in Benson and Greaves 1985, pp.50-83.

McDonald, D., 1983. 'Natural language generation as a computational problem'. In Brady and Berwick 1983, pp.209-65.

McDonald, D.D., Vaughan, M.M., and Pustejovsky, J.D., 1987. 'Factors contributing to efficiency in natural language generation'. In Kempen 1987, pp. 159-181.

Morris, P., (ed.) 1987. *Modelling cognition*. Chichester: Wiley.

Paris, C.L. 'Tailoring object descriptions to a user's expertise'. In Kobsa and Wahlster 1988.

Tench, P., 1987. *The roles of intonation in English discourse*. PhD thesis, University of Wales.

Tench, P., and Fawcett, R.P., 1988. *Specification of intonation for Prototype Generator 2*. (COMMUNAL Report No 6) Cardiff: Computational Linguistics Unit, University of Wales College of Cardiff.

Tucker, G.H., 1989. 'Natural language generation with a systemic functional grammar'. In *Laboratorio degli studi linguistici 1989/1*. Camerino: Italy: Universita degli Studi di Camerino (pp.7-27).

Zock, M., and Sabah, G., (eds) 1988a. *Advances in natural language generation Vol 1*. London: Pinter.

Zock, M., and Sabah, G., (eds) 1988b. *Advances in natural language generation Vol 2*. London: Pinter.