The parser from an Arabic text-to-speech system

Allan Ramsay, Department of Computation,
Hanady Mansour, Department of Language and Linguistics
UMIST, PO Box 88, Manchester M60 1QD, UK

Résumé - Abstract

The work described here is part of an attempt to provide a text-to-speech (TTS) system for Modern Standard Arabic (MSA). The key problem with this task is that written MSA omits a great deal of information about short vowels, and a certain amount of other phonetically relevant information, which is clearly essential for any speech synthesiser.

We have argued elsewhere that in order to recover this information you need to carry out a detailed linguistic analysis at a variety of levels. We have described elsewhere how we integrate morphological, syntactic and semantic processing. The current paper provides a detailed description of our approach to syntactic analysis of MSA. The presentation here assumes that morphological processing is interwoven with syntactic analysis, but the details of our approach to morphology are omitted: the interested reader is invited to see (Ramsay and Mansur2001).

1 Framework

The goal of our work is to provide a text-to-speech (TTS) system for Modern Standard Arabic (MSA). This is a particularly challenging task, since a great deal of the information that you need in order to produce an appropriate phonetic transcription is missing from the written form of MSA (El-Shafei2002; El-Imam2001).

Arabic speakers, of course, have very little difficulty in reading aloud. We believe that this ability is underpinned by the fact that they can mix various levels of linguistic processing, working out what the phonetic form of each word must be if the utterance as a whole is to be well-formed and meaningful.

In order to produce a TTS system, then, we need a system that can do the same task: we need a system that can take highly underspecified descriptions of sequences of Arabic words and decide what combinations of these words make well-formed, meaningful sentences.

The surface form of a word typically gives rise to a number of possible underlying forms. To take a simple example, كتب (*ktb*) could be a noun, an intransitive verb, one of two transitive verbs or their passives, or a ditransitive verb or its passive, with a range of phonetic forms. The overall goal of the work reported here is to take a sequence of written forms and obtain from them a sequence of phonetic forms which make up a syntactically well-formed, semantically plausible analysis and thence to obtain a phonetic transcription. The current paper focuses on the syntactic analysis. In practice we integrate syntactic and morphological processing very

tightly, so that we delay making choices between different nominal forms until we have determined the syntactic and semantic context (for instance we delay specifying case markers until we know whether or not something is the subject of a major clause). For the purposes of the current paper, however, we will assume that morphological processing is carried out before syntactic analysis starts, so that the function of syntactic analysis is to choose between fully determined underlying forms, rather than to help with the process of determining them.

The general framework we use for syntactic processing is as follows:

1. we are working within a sign-based framework, where a sign has five major elements:

   `structure:` this includes information about where an item appears, what words it includes, what its daughters are.

   `morphology:` this includes information about the internal structure of words. This is used largely during morphological processing, and hence we will not have much to say about it in this paper.

   `syntax:` this covers the syntactic properties of the item in question, and is clearly the most important part of the current paper. The `syntax` includes all the obvious features, but it is important to note that it also includes a list '`args`' which describes the items that are needed in order to saturate the item in question, and '`target`' which describes what this item would modify.

   `semantics:` this specifies how the meaning of the item is to be used when building an interpretation of the sentence in which it occurs. We build 'logical forms', which attempt to encode the meaning of the sentence in a form which is amenable to automated reasoning. We do this using the standard techniques of 'compositional semantics', exploiting the notion of 'glue' in order to be flexible about how an item contributes in different contexts (Dalrymple et al.1996; van Genabith and Crouch1997). We also carry a parallel shallow semantic analysis around, simply noting the semantic types of the various major constituents and using this to detect implausible readings.

   `remarks:` we use this general processing framework for a variety of applications. In some we need to record extralinguistic information, e.g. when we are using it in language learning applications, we need to record information about student errors. The `remarks` is a place for storing general extralinguistic notes.

   The example below shows a heavily edited copy of the sign for يَكْتُبُونَ (*yktbwn*), showing (i) the information that we use for constructing the phonetic transcription يَكْتُبُونَ (`yaktubūna`) (namely the morpheme sequence *'y,a,k#t#b,0, uuna'* and the selection of *'0, u'* as the diacritics to be used to fill in the gaps from the various possible diacritic sets for this form of this verb); (ii) the basic syntactic features showing that this is the third singular masculine present tense form of this verb; and (iii) the list of arguments, showing that this is a transitive form of the verb. Each entry on the argument list is in fact a complete description of a sign, saying more about what properties that this sign should have (e.g. that the first item on the list should be third singular masculine) and where it should appear (i.e. before or after the verb), but to save space we do not generally show the details of embedded signs.

   {struct(positions(start(0), end(1), span(1), +compact, xstart(0), xend(1)),
   forms({y,a,k#t#b,0,uuna}, yktbwn))),

```
morph(diacrits(choices(actvPres(["0", "u"]),actvPast(["a", "a"]),
                       psvPast(["u", "i"]),psvPres(["0", "a"])),
               actual(["0", "u"]))),
      lextype(regular(i(1, "u"), a, 1))),
syn(nonfoot(head(cat(xbar(+v, -n)),
                 agree(third(+plural)), gender(-neuter, +masculine, -feminine)),
                 vform(vfeatures(finite(+tensed, -participle, -infinitive),
                                 -aux,
                                 +active,
                                 view(tense(+present, -past, -future, -preterite, -free),
    subcat(args(["NOUN", "NOUN"]), fixed),
    foot(wh([]))),
remarks(score(0))}
```

2. For each verb we specify what kinds of arguments it expects and what their thematic roles are. From this we decide which argument should be the subject, and we specify the canonical position each argument. We use the same general technique for a range of languages, though clearly just what the canonical position is differs from language to language. We assume that for a simple verbal sentence with a transitive verb the order in MSA is VSO, though other orders are either possible or even obligatory under specified circumstances.

3. A range of word orders are possible. We exploit the parsing algorithm outlined in (Ramsay1999) to allow arguments to be found in non-canonical positions, with constraints on local subtrees to check whether such 'out of place' are in fact acceptable.

Given this general framework, we can account for a range of constructions in MSA. It is crucial to the effectiveness of the parsing algorithm that the constraints on local subtrees are evaluated on 'just-in-time' basis, so that potential analyses are ruled out as early as possible. We will point out places where this is important as they arise.

2 Verbal sentences

For verbal sentences in canonical order the situation is very straightforward. The verb specifies what the arguments should be like and where they should appear, and if there are appropriate items in the specified positions then we get an analysis (or, more often, several analyses):

(1) يكتب العالم الكتاب. (*yktb ālʿālm ālktāb.*)

(1) gives rise to numerous syntactically well-formed interpretations:

- the verb يكتب (*yktb*) has seven possible readings (as a range of active and passive forms of intransitive, transitive and ditransitive roots).
- various permutations of the canonical word order are permitted.
- the subject of a verbal sentence can be omitted, since there is enough information in the verb morphology to determine who it must have been.

We can downgrade any interpretations that assign semantically anomalous items to thematic roles, e.g. where we know that the subject of some verb ought to be a living entity, but the NP that has been assigned to this role is non-living. This leaves us with three analyses of (1), as shown in Fig. 1.

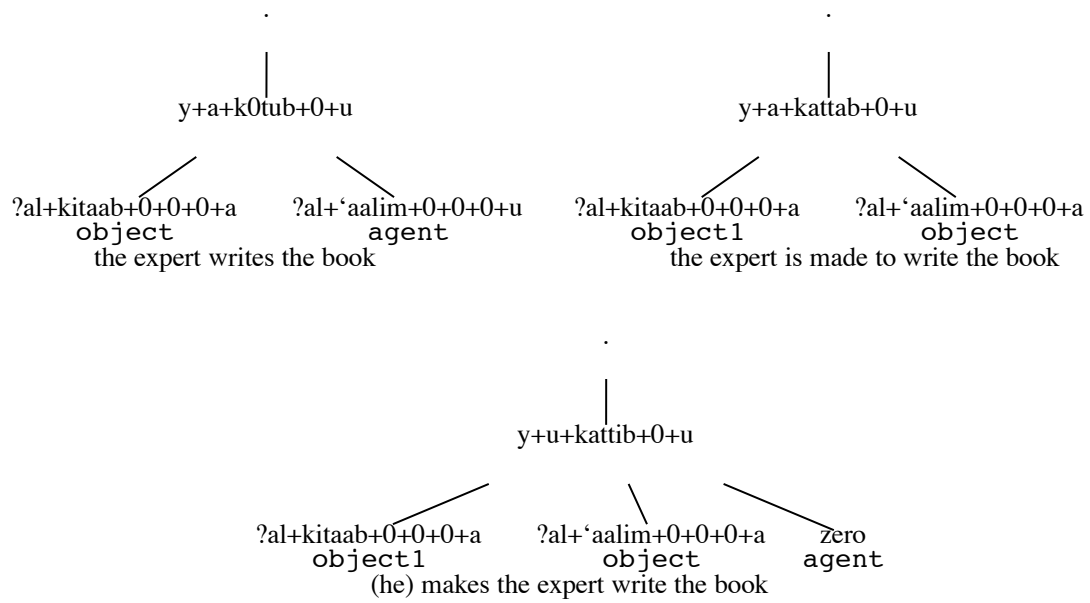

Figure 1: Interpretations of (1)

There is nothing more we can do to choose between these on syntactic or simple semantic grounds. Note that the second two, which both have كتّب (*kttb*) as the root of the verb and which both assign the same roles to the two NPs, are equivalent: the first of these employs the passive form of the verb, and hence has no visible NP denoting the agent (note the accusative case marking on the subject of the passive verb), the second is active but has a 'zero' subject for the agent, but they both convey the same proposition, so that for most purposes it doesn't matter which one you choose.

To distinguish any further, we would need to either carry out rather deep contextual reasoning or exploit statistical regularities about the cooccurrences of various forms. In fact we should probably do both, but whatever we do we need the syntactic analysis described here to show what the options are. In future work we intend to explore further the kind of contextual reasoning that might underpin making a more refined choice, but the focus in the present paper is on the basic syntactic processing.

In general, then, a written sentence of MSA may have a number of interpretations. There are, however, a number of constraints which can be exploited to reduce the level of ambiguity: the discussion below outlines a number of these constraints:

Case marking: the written form of (1) contained no visible case markers, but the analyses in Fig. 1 do include case markers on the nouns – ـُ (u) on the subject of the first analysis, since this is an active form and hence the subject should be nominative, ـَ (a) on the others (the subject of the passive verb is accusative, and all the other Naps are in object positions and hence are also accusative). We need this information for our primary application, since although the case marking is not written it is, in at least some registers, pronounced. It is worth noting, however, that in at least some cases the presence or absence of an explicit case marker can eliminate

certain readings. Consider, for instance,

(2) a. يكتب عالم الكتاب. (*yktb ʿālm ālktāb.*)

b. يكتب عالما الكتاب. (*yktb ʿālmān ālktāb.*)

(2a) has only one semantically reasonable interpretation, shown in Fig. 2.

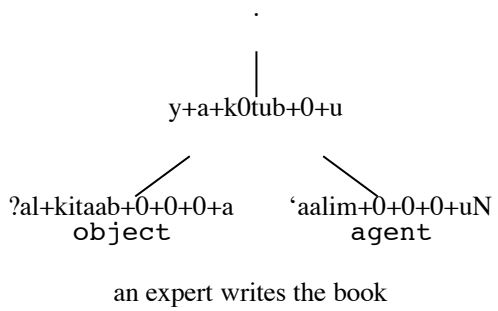

Figure 2: Lack of a case marker eliminates interpretations

The indefinite NP cannot be accusative, because the accusative form of a singular masculine NPs must have an explicit case marker اً (ān). The absence of this marker from عالم (*ʿālm*) means that this NP cannot be playing a role where it should be accusative, and hence the only reading we get is the one in Fig. 2. If, on the other hand, the accusative marker is included, as in (2b), then we only get the readings where عالما (*ʿālmān*) is not the subject of an active sentence, as in Fig. 3: the passive form is acceptable because the subject of the passive does have to be accusative, and the form with the zero subject is also acceptable because then عالما (*ʿālimān*) is the object and hence is accusative.

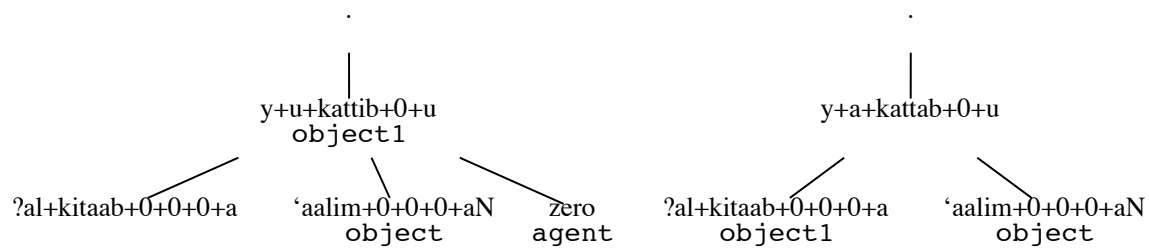

Figure 3: With the case marker we get different possibilities

.

Agreement and Order:

Arabic has a complex set of agreement markers, covering both number and gender, and there is a very general constraint that the verb and the subject should agree in gender which can be used to rule out potential analyses. It also allows a number of permutations of word order, usually for discourse reasons, but there is a particularly interesting constraint on agreement in non-canonical orders. In particular, Arabic sentences can have SVO order, but only if the subject is definite, and under these circumstances the subject must agree in gender *and number* with the verb. Consider, for instance, (3)

(3) a. الدرس كاتب المدرس. (*āldrs kātb ālmdrs.*)

b. كاتب الدارسان المدرسان. (*kātb āldārsān ālmdrsān.*)

c. الدارسان كاتب المدرسان. (*āldārsān kātb ālmdrsān.*)

We get two readings for (3a), one with الد س (*āldrs*) as subject (SVO) and one with الد س (*ālmdrs*) as subject (OVS). They are both marked, since they both deviate from the canonical order, but they are both possible.

The situation in the (b) and (c) examples is more complex. In both these examples the verb is third singular, both the nouns are third dual. For (3b) we get one reading with الدارسان (*āldārsā-n*) as the subject and another with المدرسان (*ālmdrsān*) as the subject, since the only constraint on the subject is that it should agree in gender with the verb. For (3c), however, we only get one reading. The SVO reading is ruled out because it would involve left-shifting of the subject, and if that happens then the subject and the verb must agree completely, not just in gender.

3 Nominal sentences

In addition to sentences where there is a main verb and a set of arguments, Arabic allows 'nominal sentences', consisting of a subject NP and a predication (i.e. a second NP, an adjective or a PP). Intuitively, these sentences have a meaning very like an English copula sentence, but they do not contain a copula.

It is very tempting to try to deal with these cases by 'hallucinating' a copula, but it is very hard to see how to implement this within a lexical grammar of the kind outlined above. The general grammatical tradition we are using involves complex lexical items which contain detailed information about the relations that the item is prepared to enter into, together with a skeletal set of rules that lay out the general principles of combination. Within such a framework, it seems inappropriate to insert invisible lexical heads. We prefer, therefore, to exploit a distinction between the 'internal' and 'external' views of an item.

The notion we are using here is illustrated by English gerunds. Consider (4):

(4) He concluded the banquet by eating the owl.

'eating the owl' here looks like a VP. It contains a verb, *'eating'*, and a direct object NP *'the owl'*. Note that it is unlikely that *'eating'* in this example is a noun, since if it were then you would expect a determiner and would also expect the complement to be marked by *'of'*, as in

(5) The RSPCA disapproves of the eating of owls.

We deal with verbal gerunds of the kind shown in (4) by introducing 'post-lexical rules', similar to (Nerbonne et al.1994)'s use of 'lexical rules' for dealing with nominal gerunds of the kind illustrated in (5). Such rules look a bit like phrase structure rules, but they always have exactly one item on the right. The rules for the two cases are roughly as follow:

```
% nominal gerund
noun[args=A] ==> verb[+prespart, args=A]

% verbal gerund
NP ==> VP[+prespart]
```

Each of these rules that something which looks like an item of the kind described on the right can be used in contexts where something of the kind on the left is required. This enables us to allow the head of the nominal gerund to behave like an ordinary noun (so it combines with determiners, and its object is case-marked by *'of'*), whereas the head of the verbal gerund

behaves like a verb (no determiner expected, object not marked by *'of'*) but the whole thing can be used in contexts where an NP is required.

We exploit this general idea by introducing a post-lexical rule for Arabic which says that an NP can be regarded as the subject of an incomplete nominal sentence which still requires a predication.

```
S(args=[PRED]) ==> NP
```

This says that if you have an NP, then if you find something predicative in the right place then you can combine them to make a sentence. The details of what the predicative item should be like, and where you expect to find it, are fairly complex. The full form of the rule we use has the following constraints added:

- The canonical order specifies that the predication should follow the subject.
- The predication should either be nominative or marked by a preposition.
- If the subject is definite then the standard order should be maintained.
- If the subject is indefinite then the predication should be marked by a preposition, and the canonical order should be reversed.

Given these rules, we obtain the analyses in Fig. 4 for (6).

(6) a. الدارس عالم. (*āldārs ʕālm.*) (the student is an expert)

b. في المكتب كتاب. (*fy almktb ktāb.*) (a book is in the office)

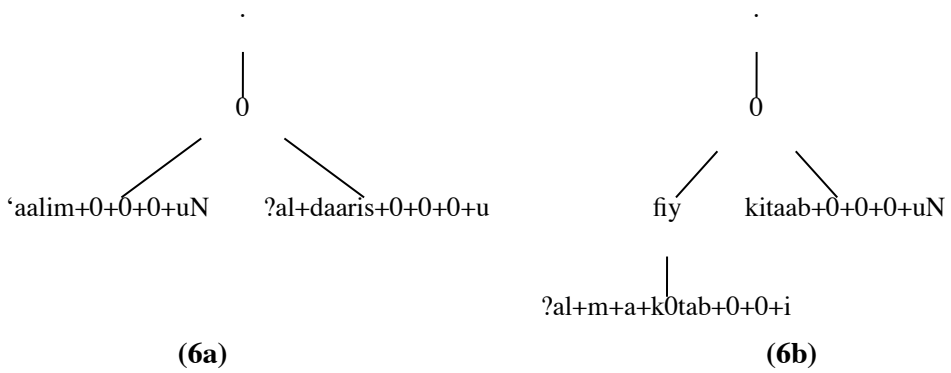

Figure 4: Nominal sentences

In (6a), عالم (*ʕālm*) cannot be the topic, since it is indefinite and hence it would have to be the first item in the sentence if it were the subject. Hence عالم (*ʕālm*) and الكتاب (*ālktāb*) are both nominative, الكتاب (*ālktāb*) because it is the subject and عالم (*ʕālm*) because the predication must be either nominative or marked by a preposition. In (6b), في الكتاب (*fy ālktāb*) cannot be the topic, since it is not nominative. So عالم (*ʕālm*) becomes the subject, which is OK because it is

in initial position, and hence again gets a nominative case-marker, whereas الكتاب (*ālktāb*) gets the genitive marker because it is the complement of في (*fy*).

The constraints on order and case marking outlined above again mean that although any pair of adjacent NPs potentially make a nominal sentence, with some scope for ambiguity, in a very large number of cases they either cannot combine or can combine in only one way.

4 Construct NPs

Adjacent nouns can also combine to form 'construct NPs', where roughly speaking the second NP in the pair is a possessive determiner. The possessive NP must be definite, and must follow the target noun:

(7) كتاب العالم (*ktāb ālʕālm*) (the book of the expert, the expert's book)

The interpretation is given in Fig. 5. It is notable that the nucleus gets assigned the nominative case marker for a *definite* nominative NP. The underlying reason for this may be seen in the parallel with the English construction *'the expert's book'*, where the definite article combines with *'expert'* to pick out a specific expert, and then the whole NP refers to the book belonging to that expert. So the whole NP is definite, even though there is no explicit definite article for *'book'*. Similarly the whole NP in (7) is definite, despite the lack of a definite article on كتاب (*ktāb*), and hence the case marker is the one that is attached to definite NPs[1].

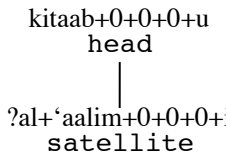

Figure 5: The expert's book

Note that the presence of construct NPs like this undermines the claim above that (2a) only has one interpretation that satisfies the selection restrictions on the verb, since we now also get the interpretation in Fig. 6, with the construct NP as the subject of an intransitive reading of يكتب (*yktb*). As before, choosing between this reading and the one in Fig. 2 requires deeper reasoning than can be done just using selection restrictions.

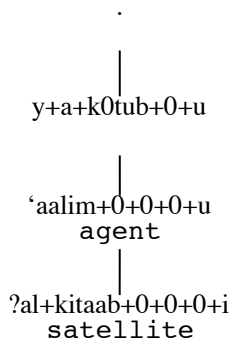

Figure 6: The book's expert writes

[1]Why it should always be the nominative form is less easy to explain. Note, however, that the presence of the nominative case marker on the head noun does not mean that NP as a whole is nominative.

5 Clausal complements

We note finally that simple sentences may involve verbs with clausal complements. We deal with these by allowing the embedding verb to place constraints on the form of the embedded sentence (e.g. whether it should be indicative or subjunctive or jussive) and on whether or not it must/may/must not contain a complementiser. As an example, the verb اعتقد (*ā؟tqd*) (thought) requires a complement headed by the complementiser خنّ (*ẖann*), where the complement is tensed/indicative, is in canonical order, and has an accusative subject. This produces the analysis in Fig. 7 for the example in (8). Other matrix verbs place different constraints on the embedded argument, but for most such verbs it is a simple matter of specifying what the form of the embedded sentence should be and whether it should have a complementiser, with a tight link also between the chosen complementiser and the form of the embedded clause (Mohammed2000).

(8) اعتقد خنّ الكتاب في المكتب. (*ā؟tqd ẖann ālktāb fy ālmktb.*) (he thought that the book (was) in the office.)

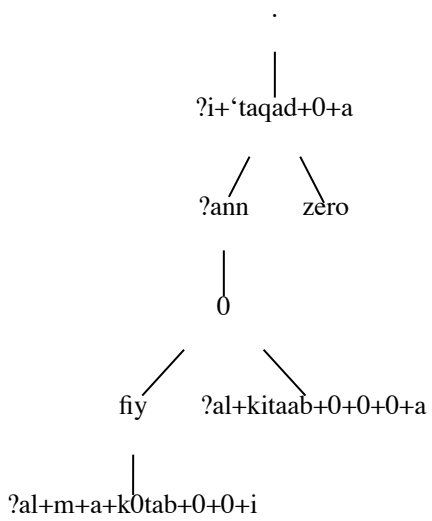

Figure 7: Analysis of (8)

6 Conclusions

The parser outlined above was developed as part of a text-to-speech system for MSA. Within this application the parser is used in tight connection with the morphological analyser, with the morphological analyser proposing underspecified descriptions of lexical items which are input to the parser and which get further refined as the syntactic analysis proceeds. It is therefore important for us to recover the selection of appropriate diacritics, case markers, agreement markers and other items that are omitted in the written form. The analyses given above show that this is dealt with satisfactorily, at least for the kind of rather simple examples in the current paper.

The system’s performance does degrade as the input texts get longer. The time taken to carry out morphological and syntactic analysis of (8), for instance, is around 0.65 seconds, which is perfectly acceptable for our task of reading text aloud (a TTS system which takes about the

same amount of time to process the text as it takes to actually pronounce it will sound reasonably fluent after a slight initial delay). As the input texts get longer they get more ambiguous, which means that the analysis takes longer, but more importantly it also means that we can be less confident that we know which is the right one. We can sidestep this to some degree in the TTS application, since many of the analyses will have either identical or very similar phonetic transcriptions, so that it doesn't actually matter very much which one we choose. For other applications this will become increasingly significant. It seems unlikely that simple selection restrictions will prove powerful enough to help us make the right decision in complex situations. We are therefore planning to exploit a slightly richer set of rules, couched in a simple description logic, in order to reason more accurately about which interpretations are most plausible.

The system has been linked up to the MBROLA speech synthesiser(Dutoit et al.1996; Dutoit1997), and does produce comprehensible spoken output. MBROLA allows you to specify pitch contours, and we are trying to use general rules about stress in Arabic to produce more natural sounding output, but this work is also still under investigation.

References

M Dalrymple, J Lamping, F C N Pereira, and V Saraswat. 1996. A deductive account of quantification in LFG. In M Kanazawa, C Piñón, and H de Swart, editors, *Quantifiers, deduction and context*, pages 33–58.

T Dutoit, V Pagel, N Pierret, F Bataille, and O van der VrekenVAN DER VREKEN. 1996. The mbrola project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In *Proc. ICSLP'96, vol 3*, pages 1393–1396, Philadelphia.

T Dutoit. 1997. *An Introduction to Text-To-Speech Synthesis*. Kluwer Academic Publishers, Dordrecht.

M A El-Imam. 2001. Synthesis of Arabic from short sound clusters. *Computers, speech and language*, 15:355–380.

M El-Shafei. 2002. Techniques for high quality Arabic speech synthesis. *Information Sciences*, pages 255–267.

A Mohammed. 2000. Word order, agreement and pronominalisation in standard and Palestinian Arabic. *Current Issues in Linguistic Theory*, pages 1–81.

J Nerbonne, K Netter, and C Pollard, editors. 1994. *German in Head-Driven Phrase Structure Grammar*. CSLI Lecture Notes, Center for the Study of Language and Information, Stanford.

A M Ramsay and H Mansur. 2001. Arabic morphology: a categorial approach. In *ACL workshop on 'Arabic language Processing: Status and Prospects'*, pages 17–22, Toulouse. Association for Computational Linguistics.

A M Ramsay. 1999. Direct parsing with discontinuous phrases. *Natural Language Engineering*, 5(3):271–300.

J van Genabith and R Crouch. 1997. How to glue a donkey to an f-structure. In H Bunt, L Kievit, R Muskens, and M Verlinden, editors, *2nd International Workshop on Computational Semantics*, pages 52–65, University of Tilburg.