

- Lehmann, W. 1973. 'A structural principle of language and its implications'. *Language* 49, 47-66.
- Li, Charles N. & Sandra A. Thompson. 1976a. 'Subject and topic: a new typology of language'. In Li & Thompson 1976b, 458-89.
- Li, Charles N. & Sandra A. Thompson (eds.). 1976b. *Subject and topic*. New York: Academic Press.
- Lindvall, Ann. 1998. *Transitivity in discourse*. Lund: Lund University Press.
- Matisoff, James A. 1973. *The grammar of Lahu*. Berkeley: University of California Press.
- Molnár, Valéria. 1998. 'Topic in focus: on the syntax, phonology, semantics and pragmatics of the so-called "contrastive topic" in Hungarian and German'. *Acta linguistica hungarica* 45, 1-68.
- Prince, Ellen. 1981. 'Toward a taxonomy of given-new information'. In Peter Cole (ed.). *Radical pragmatics*, 223-255. New York: Academic Press.
- Rosén, Victoria. 1998. *Topics and empty pronouns in Vietnamese*. Diss., University of Bergen.
- Schachter, Paul. 1976. 'The subject in Philippine languages: topic, actor, actor-topic or none of the above'. In Li & Thompson 1976b, 493-518.
- Schachter, Paul. 1985. 'Parts-of-speech systems'. In Timothy Shopen (ed.), *Language typology and syntactic description* I, 3-61. Cambridge: Cambridge University Press.
- Shibatani, Masayoshi (ed.). 1988. *Passive and voice*. Amsterdam: Benjamins.
- Svantesson, Jan-Olof. 1994. 'Kinesiska'. In Vamling & Svantesson 1994b, 197-209.
- Vallduví, Enric & Elisabet Engdahl. 1996. 'The linguistic realisation of information packaging'. *Linguistics* 34, 459-519.
- Vamling, Karina & Jan-Olof Svantesson. 1994a. 'Inledning till språktypologin'. In Vamling & Svantesson 1994b, 1-30.
- Vamling, Karina & Jan-Olof Svantesson (eds.). 1994b. *Världens språk: en typologisk och geografisk översikt*. Dept. of Linguistics, Lund University.
- Venneman, Theo. 1972. 'Analogy in generative grammar, the origin of word order'. In L. Heilman (ed.). *Proceedings of the Eleventh international congress of linguists* 2, 79-83. Bologna: Il Mulino.

Ann Lindvall, IMER, Malmö högskola <Ann.Lindvall@imer.mah.se>

Hit and *dit* in translations between Swedish and English

Ingrid V. Nilsson

1 Introduction

1.1 Problem

Translating between two or more languages, the question of the translative validity of corresponding or equivalent words in different languages is eventually raised (Jakobson 1959), even between languages as similar in structure as Swedish and English. Swedish situative adverbs/particles, especially in their directive, or lative, function (*hit/dit*, etc.), pose such a problem – especially when translated to or from a language without overt manifestation of this functional contrast. In relation to the speaker (or the point of origin of the statement/story), Swedish can show location as well as direction; situatedness close to, as well as distant from, and motion towards as well as away from, the deictic center. English, on the other hand, mostly uses locational expressions (*here/there*), and also employs fewer such situative words than does Swedish (Jespersen 1964). (For an illustration of English verb complements, see Svartvik & Sager 1996.)

Furthermore, these Swedish adverbs/particles ('deictic situatives') may, or may not, be part of a verb phrase (*ta sig dit*), combined with a verb (*komma hit*), and/or with another adverb/particle included (*cykla upp hit*). Their grammatical functions can be those of being purely an adverb, or part of a phrasal verb, or being used with extra – added – adverbs with or without a phrasal verb, or as relative adverbs initiating a relative clause (Thorell 1973, Teleman 1974). Of these functions, some uses may overlap and some can fill more than one function.

2 Issues to be looked at in this study

2.1 Deictic situatives, and the ways they can be approached

Locative/directive adverbs or particles either indicate an 'essive' (stative) location of something in relation to the speaker/point of view (close/remote), or

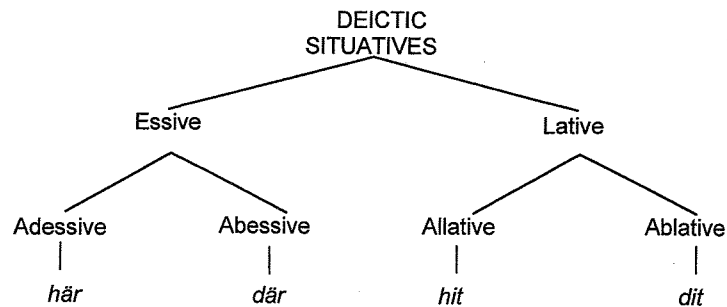


Figure 1. Deictic situatives in Swedish

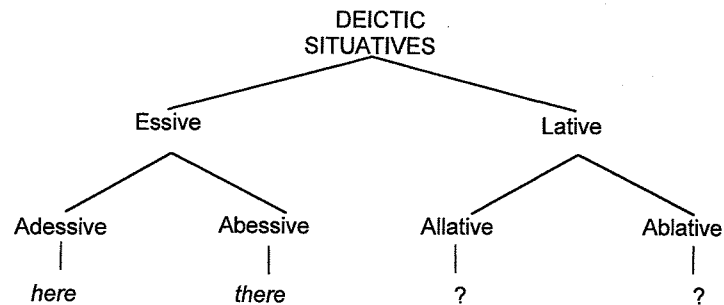


Figure 2. Deictic situatives in English

they indicate a 'lative' (dynamic) motion of something in relation to the speaker/point of view (approaching/distancing), or – when combined – a 'distributive' characteristic, which can be either 'essive' (*här och där*), or 'lative' (*hit och dit*). In an adaptation of the so-called Comrie notation (Comrie 1981, as adapted in Strömquist & Verhoeven 2004), *hit* could illustrate a deictic function with an 'allative' (towards the speaker/point of view) direction, and *dit* a deictic function with an 'ablative' (away from the speaker/point of view) direction; *här* and *där* could signal deictic functions expressing an 'adessive' (close to the speaker/point of view) and an 'abessive' (distant from the speaker/point of view) location, respectively. An illustration of these concepts is given in Figures 1 (Swedish) and 2 (English).

In terms of finding straightforward correspondences, essive expressions should pose fewer problems for translators than lative, because of the close

similarity in meaning between Swedish *här/där* and English *here/there*. Examining the translations of *hit/dit* into and from English, however, could give evidence of either individual or systematic solutions to the problem of how to render the expressly lative quality of the Swedish *hit/dit*, or of how to insert this quality into a Swedish translated text. Variations in translational strategies could then either be based on language specific means of expression – or lack thereof, or on other common strategies, i.e. ignoring the deictic lative quality completely, or rephrasing it. If so, perhaps other language specific means of evidencing deictic lative qualities could be pinpointed.

2.2 Hypotheses about translative strategies

2.2.1 Hit/dit: Swedish to English.

- (1) Literal translation equivalents: Swedish lative words = English essive words: Swedish *hit och dit* = English *here and there*; situative, but not lative.
- (2) Varying translations: Swedish lative words = English situative non-literal replacement translations: Swedish *hit* = English *to me, to this place, etc.*
- (3) Without translation: Swedish lative words = English \emptyset : Swedish *dit* = English \emptyset , e.g. *De gick dit för att hämta ... = They fetched ...*

2.2.2 English to Swedish. Strategies for translating from English would presumably show similar patterns:

- (1) Essive words translated with their Swedish essive equivalents
- (2) Lative words being inserted instead of essive words when appropriate for Swedish usage, but also
- (3) Lative words inserted where there are no equivalents in English.

2.3 Specific questions to be examined

This study will examine how the Swedish words *hit/dit* (and their distributive combination: *hit och dit* (+ variations)) have been used in translations to and from Swedish and what their corresponding English expressions are, as seen through a number of shorter original Swedish texts translated into English, and vice versa, by various translators. The corpus, from which the samples are gathered, is the *English Swedish Parallel Corpus*, structured and compiled mainly by Karin Aijmer, University of Gothenburg, Bengt Altenberg, and Mikael Svensson, University of Lund.

Also interesting to see is whether there are any eventual systematic tendencies when insertions of *hit/dit* into Swedish translations have taken

place. When such insertions are made, are *hit/dit* used as language specific alternatives for *here* and *there*, or are they without equivalents in the English version? If they do lack straightforward equivalents, are they then a means of translating situative functions of the lative type indicated through other words or expressions in the English version, or is it simply a lexical difference? If it is not a purely lexical difference – how are such lative properties manifested? Is it possible for a lative characteristic to be ‘floating’, i.e. shown through other words or functions in a sentence?

A third question of interest is a comparison of the lexical/grammatical functions in the sentences of *hit/dit*. When evidenced by separately translated words – i.e. not incorporated into another word, such as a verb – do they, for instance, most often form part of a verb-particle phrase, or do they mainly indicate direction (adverbial), or is their prevalent function that of starting an adverbial relative clause? (Thorell 1973). In the latter case, a correspondence between *där/dit* and *where* could furthermore be expected (Svartvik & Sager 1996).

3 Method

3.1 Corpus from which the material for the study is sampled

In the bi-directional *English-Swedish Parallel Corpus* (second, enlarged, version completed in 2001, by Altenberg, Aijmer and Svensson), written modern fiction and non-fiction, translated as well as original, have been recorded, and cross-referenced in a corpus made available to researchers. British English, as well as American English, is represented, and also Australian, Canadian, and Irish English. It consists of a representative selection of fiction, as well as non-fiction. No spoken language is represented, nor newspaper texts or correspondence of any kind. There are 50 original texts in the fiction section of the corpus evenly divided between Swedish and English, (25 Swedish, 25 English), and 86 original texts in the non-fiction section of the corpus (47 Swedish, 39 English). The size and the proportion of the texts used, however, are roughly equal between both genres and languages. The majority of all texts are published after 1980. The current approximate size of the corpus is 2.8 million words. (See the *References* for links to the *ESP Corpus*, and to the information about authors, translators, and texts used in the corpus.)

Table 1. Possible search inputs and outputs

<i>Search input</i>	<i>Parallel output</i>
Swedish fiction original	English fiction translation
Swedish non-fiction original	English non-fiction translation
Swedish fiction translation	English fiction original
Swedish non-fiction translation	English non-fiction original
English fiction original	Swedish fiction translation
English non-fiction original	Swedish non-fiction translation
English fiction translation	Swedish fiction original
English non-fiction translation	Swedish non-fiction original

3.2 Conditions on searches

The *ESPC* is a bi-directional translation corpus. As such, it can be searched with reference both to the original language of the texts used, and to its translation.

The search is performed as a sentence search, and it is annotated with regard to position of the sentence in the text alternative searched. Its equivalent – and position – in the parallel text is similarly annotated and listed, a T indicating a translated version. Depending on which version is of interest, either translation or original can be presented first.

The search type is a version of a KWIC search – showing the keyword in context – and the length of the context string is mostly defined as the orthographic sentence. Any discrepancy between the languages is resolved according to the language with the longest sentence. If the parallel language has two sentences for the translation, both of these are given – in most cases.

This means that there are four choices for each language, in terms of a primary search, for which the other language's sentence-unit version(s) is simultaneously shown (See Table 1). In terms of this study, only the alternatives for Swedish searches were used, since the item of primary interest was the use of the Swedish words *hit/dit* in originals as well as in translations.

3.3 Category groupings of search results

There are two basic groupings for *hit/dit* used in this study:

- According to semantic-syntactic function in the sentence
- According to English counterpart, or translation type

Both of these categories are calculated separately, as well as combinatorially, for *hit* and *dit*. The classification according to function combines grammatical and lexical functions, and is grouped into instances that are experienced to be:

- Strongly linked to a verb
- Weakly linked to verb, with more of a lative function (allative: *hit*, or ablative: *dit*)
- Relative adverb (*where*)
- Distributive function (*hit och dit*, and variants thereof)
- Other functions, especially idioms (see Table 3)

The division between the groups 'strongly' and 'weakly' linked to a verb is partly lexical (what is specified in a dictionary as a commonly occurring verb+particle combination containing *hit/dit* under the main entry for the verb in question), partly semantical (what specific function *hit/dit* fills in that particular sentence). If *hit* or *dit* could be removed from the sentence without altering more than the directional indications, it was deemed to have a primarily lative function and to be weakly linked to a verb, e.g. *förflytta sig dit*, where the basic meaning of the verb remains the same and *dit* supplies the directional input. If – upon removal of *hit* or *dit* – the verb changed meaning, or the sentence became incomplete, it was deemed to be strongly linked to the verb, e.g. *höra hit*, where the verb without *hit* receives a completely different interpretation. This group of verbs + *hit/dit* furthermore often forms participles with *hit/dit* included, e.g. *ditresande*, etc. (Teleman 1970).

The translation types are grouped according to the questions posed in the hypotheses about translation strategies (see Table 4), i.e.;

- From/into *here*, or *there*, or *where*
- From/into other equivalent
- Without equivalent

4 Findings

There were 447 instances found of *hit/dit*, 182 of *hit*, and 265 of *dit* (see Table 2). The higher number of instances of *dit* depended partially on the relative adverb-function of this word in Swedish – 51 examples of such a function were found (see Table 3) – and partially on its more extensive use in a 'strong' verb combination together with a large variety of verbs other than *komma* (104 examples, versus *hit*, in a 'strong' verb combination, in only 42

Table 2. The number of instances and percentages of search word found in *ESPC* searches (Search language is Swedish. F = fiction, NF = non-fiction; T = translation, O = original.)

Search word	No/text type	<i>hit/dit</i> in % of no. of words/ text type
<i>hit</i> NF T	16	0.00464
<i>hit</i> F T	70	0.02019
<i>dit</i> NF T	33	0.00958
<i>dit</i> F T	84	0.02423
<i>hit</i> NF O	18	0.00509
<i>hit</i> F O	78	0.02531
<i>dit</i> NF O	51	0.01443
<i>dit</i> F O	97	0.03147
Total	447	

examples, see Table 3). In 114 instances (81 *hit*, 33 *dit*) the use of these words was tied to the verb *komma*, indicating that *komma hit* can be treated as one lemma, but *komma dit* maybe cannot.

There were 104 instances of *hit/dit* used in a lative capacity; *hit* in an allative function (40), and *dit* in an ablative function (63); see Table 3.

Hit och dit and the various versions of this 'distributive' expression, were found in 13 instances (although one such instance was not listed in the *ESPC* under the search concerning *dit*). For the instances found of *hit och dit*, etc., type of text (fiction/non-fiction) was not significant – which was an unexpected result, since (according to the author's intuitions) this expression is perhaps felt to belong in spoken or fiction-type language use more than in written or non-fiction such.

Table 3. Instances of *hit/dit* by syntactic-semantic function in Swedish. X denotes a verb other than *komma*. (The discrepancy between the two distributive numbers for *hit (och) dit* is due to a missing translation in the *ESPC* for one instance.)

	<i>hit</i>	<i>dit</i>	total
<i>kom</i> * _	81	33	114
X* _	42	104	146
allative	40		40
ablative		63	63
relative adverb	1	51	52
distributive	13	12	25
other	5	2	7
total	182	265	447

Table 4. Total number of occurrences of *hit/dit* (to and from Swedish) and types of translations found

Translations	Number
<i>here – hit</i>	91
<i>there – dit</i>	72
<i>where – dit</i>	24
other	132
without equivalent	128
Total	447

Seven instances of *hit/dit* that were termed 'other' were found, mostly idiomatic uses of *hit* or *dit*, which qualified neither as verb particles nor as latives according to the criteria applied in this study. Three of these were instances of *hit men inte längre*, which was deemed an idiom, more than a true lative situative.)

There were quite a few instances found of use/insertion of *hit/dit* without corresponding English expressions (128 examples out of 447 total occurrences, see Table 4). Of the examples with some kind of translated situative correspondence the following results were found: There were 91 literal correspondences *hit/here*, 72 literal correspondences *dit/there*, 24 correspondences *dit/where*, and there were 132 instances of 'other' translations of *hit/dit*.

The guidelines used for evaluating how the translations were done were quite literal. If the words *hit/dit* and the words *here/there* occur in the sentence unit in a one-to-one relationship – regardless of other linguistic surroundings – it was deemed a literal correspondence. (E.g. ..., *och komma iväg dit!...*, and to go *there*.)

If there were other words found (prepositions, idiomatic expressions, etc.) that could be regarded as corresponding to the directional implications of *hit/dit*, it was counted as an occurrence of 'other' (132 instances out of the total of 449). (E.g. *När jag reser hit .../Travelling to this place ...*) (For a discussion of the cognitive spatial implications of English prepositions, see Herskovits 1986.) But if there were no actual words corresponding to the directional implications of the words *hit/dit*, it was counted as an instance of a language specific use/insertion without English equivalent. (E.g. ... *och så kunde A. sätta dit en liten .../and then A was able to add a small ...*) This classification can undoubtedly be improved with further studies.

This part of the study does not cover the English equivalents and translations of *here/there*, and so any atypical translations or insertions of these

words, and/or occurrences without equivalents in the Swedish texts, will have to await further analysis.

5 Comparison

5.1 Swedish to English

In terms of the hypotheses about translative strategies into English for *hit/dit*, the following observations can be made:

1. Swedish lative words \Rightarrow English situative words (essive, or other replacements)

The general assumption that there would be various kinds of translations, many with some sort of situative connection, proved to be true for about 71% of the cases of *hit/dit*, and for nearly 75% if the relative directional adverb *dit* = *where* is included.

2. Swedish lative words \Rightarrow English \emptyset

This hypothesis proved true for almost 25% of the cases, where there were no equivalents at all to the Swedish *hit/dit*. No other part of the sentence, according to the parameters specified, could be defined as having assumed the lative function.

The 75% of cases where some kind of situative translation was found can further be divided into:

- a. Swedish lative words = English essive words

In almost 39% of these cases *here/there* corresponded to *hit/dit*, i.e., a situative rendering exists, but without the lative characteristic, showing only the essive part of the situative function.

- b. Swedish lative words = English XXX (other part of sentence)

In approximately 31% of the English versions of *hit/dit* the lative function was taken over by other expressions in the sentence – often verbs such as *arrive/depart*, *insert/remove* that already contain a directional characteristic, but also by other situative expressions such as a repetition of the actual designation for the place in question (*into the room/to Stockholm/to the hospital/ etc.*) – leaving the Swedish words *hit/dit* without 'literal' translation.

Overall, it could be established that *dit* had more instances of 'other' (36%) or \emptyset (29%) translations than *hit* (24% 'other', 19% \emptyset) in the texts translated

into English. This may be due to the numerous instances of *komma hit*, which is generally translated into *come here*, and where the conformity between Swedish and English is much stronger than in other verb-particle combinations.

5.2 English to Swedish

Concerning the translation into Swedish, the hypotheses for *hit/dit* were similar:

- a. Lative words being inserted instead of essive words when appropriate for Swedish usage, but also
- b. Lative words inserted where there are no equivalents in English.

In around 32% of the cases *here/there* were translated into *hit/dit*. Together with *where-dit* (6%) this type of translation represented 39%. In 27% of the instances of *hit/dit*, these were substituted for other situative expressions in English, and in 33% of the cases they were inserted where no English equivalents were found. Again, in the translation to Swedish, *dit* had a higher percentage 'other' translations than did *hit* (30% vs. 23%) – presumably for the same reason as in the translations into English (i.e. *come here – komma hit*), but only a slightly higher percentage of \emptyset translations (34% vs. 33%).

6 Discussion

It is obvious that there is a strong language-specific characteristic in the Swedish use of *hit/dit* that does not have an equivalent in English. Both the orientation and the direction of the motion of something or someone in relation to the center of focus in the sentence is an important trait in Swedish descriptions, whereas mainly the orientation and the process (as depicted mostly through the verb/s) seem to be of interest in an English description.

6.1 Interpretations

The results can be interpreted according to three different criteria:

- There are mostly language specific translations of the Swedish situative lative function (56% into English, and 60% into Swedish), as illustrated through the use/s and translation/s of the Swedish *hit/dit*.
- The total situative function is comparable between Swedish and English, as shown in the conformity between translations English to Swedish, and Swedish to English, if the distinction essive versus lative is left out.

- There might be a different characteristic with which a situative/lative function can be evidenced through other parts of the sentence.

Should the translations of the situative function be purely language specific, there ought to be no agreement in *hit/dit* translations. And, the results show that of 182 instances of *hit*, there are 91 direct equivalents *hit-here*, and of 265 instances of *dit*, there are 96 direct equivalents *dit-there-where*. This result demonstrates an overlap of around 43% between situative translation equivalents, if the lative – essive distinction is ignored, and only the basic characteristic 'situative' is calculated.

The 'other' renderings of *hit/dit* (i.e. situative/lative quality maintained, even though not literally translated) covers 132 cases out of 447 (30%). Expressions without equivalents cover 128 cases out of 447 (29%). Together, these two types of translation make up 58% of the total inventory of *hit/dit*. This means that there is a fairly strong difference between English and Swedish in how lative qualities are rendered. For two languages as much alike as Swedish and English historically are, there ought to be a rather large correspondence between the translations of deictic situative expressions. That is, however, obviously not the case. The number of 'non-equivalents' is too high, and the number of relative equivalents is too low.

Should there be another way of demonstrating lative qualities in English, this has to be sought through some kind of floating characteristic, designating a space/direction quality, which may be alternatively attached to different parts of a sentence, and the use of which may be more metaphorical than lexical.

7 Conclusions based on the ESPC Material

Specific:

- For translations into English, the correspondence of *hit/dit* to *here/there-where* is relatively high – around 43%.
- For translations from English *here/there-where*, the results are somewhat lower (around 39%).
- For translations to or from English, the 'non-translation', but 'circum-translation', of expressions covering the lative function is comparable – 31% into English, and 27% into Swedish.
- For translations evidencing insertions/omissions of *hit/dit*, the instances are higher for English to Swedish (33%) than for Swedish to English (25%).

General:

- Expressions of directionality may have different information-carrying roles in different languages.
- This role, as perceived by the translator, is conveyed through both position and arrangement of such expressions in the translation.
- Directionality is a universal characteristic of language; ways to express this are not.
- Directionality can be evidenced in words or phrases with other unrelated semantic-syntactic functions.
- Deixis is an important factor when a decision is taken whether to include directionality or not.
- According to the markedness theory, location/stasis should be unmarked, and direction/dynamism marked. Consequently, not all languages may have the function-specific words to express directionality, but will evidence other means for expressing this quality.

References

- Comrie, Bernard. 1981. *Language universals and linguistic typology*. Chicago: University of Chicago Press.
- Herskovits, Annette. 1986. *Language and spatial cognition*. Worcester: Cambridge University Press.
- Jakobson, Roman. 1959. 'On linguistic aspects of translation'. In Reuben A. Brower (ed.), *On translation*, 232-39. Cambridge, Mass.: Harvard University Press.
- Jespersen, Otto. 1964. *Essentials of English grammar*. Forge Village: University of Alabama Press.
- Strömquist, Sven & Ludo Verhoeven (eds.). 2004. *Relating events in narrative: typological and contextual perspectives*. Mahwah: Lawrence Erlbaum Associates.
- Svartvik, Jan & Olof Sager. 1996. *Engelsk universitetsgrammatik*. 2nd ed. Stockholm: Almqvist & Wiksell.
- Teleman, Ulf. 1970. *Om svenska ord*. Lund: Gleerups.
- Teleman, Ulf. 1974. *Manual för grammatisk beskrivning av talad och skriven svenska*. Lund: Studentlitteratur.
- Thorell, Olof. 1973. *Svensk grammatik*. Stockholm: P.A. Norstedt & Söner.
- Link to the opening page of *ESPC*: <http://www.englund.lu.se/research/corpus/access.phtml>
- Link to the list of texts included in the *ESPC*: <http://www.englund.lu.se/research/corpus/corpus/webtexts.html>

Ingrid V. Nilsson <Ingrid.Nilsson@ling.lu.se>

Automatic estimation of speaker age using CART

Susanne Schötz

This paper describes a small attempt to automatically estimate speaker age aimed at increasing the phonetic knowledge of age. Acoustic features were extracted from the four phonemes of the Swedish word /ra:sa/ 'collapse' produced by 428 adult Swedish speakers, and then used to build CARTs (Classification and Regression Trees) for prediction of age, age group and gender. Results showed that the CARTs used different strategies to estimate different phonemes, and that age predictors for /a:/ and /s/ performed best. The best CARTs made about 91% correct judgements for gender, about 72% for age group, while the correlation between biological and predicted age was about 0.45. When comparing these results to those of a previous study of human age perception, it was found that although humans and CARTs used similar cues, the human listeners were somewhat better at estimating age. More studies with a larger and more varied speech material are needed in further pursuit of a good automatic age estimator.

1. Introduction

Verbal human-computer communication distinguishes itself from human-to-human communication in many ways. One difference is that most systems fail to identify the speaker-specific or paralinguistic information present in every voice. Human listeners almost instantly recognize the gender, emotional state, attitude and state of health of a speaker. Even age is fairly well judged by listeners. If human-computer interfaces were able to capture some of these properties, man-machine communication would become more natural. Spoken dialog systems would be able to adapt to the gender, age and other speaker characteristics of the user, which could lead to increasing performance. This paper describes a small attempt to automatically predict one speaker-specific quality: age, using an important technique in pattern recognition: CART, and then comparing the results to age judgements of human listeners.

1.1 Background

While researchers agree that human listeners are able to judge speaker age to within ± 10 years, few computers have had a go at this task. One reason for this may be that it is far from easy. There are acoustic correlates to age in