The Fyntour Multilingual Weather and Sea Dialogue System - Description and Assessment

Jens Ahlmann Hansen[a], Eckhard Bick[b*]

[a]Dept.of Business Communication and Information Science, University of Southern Denmark
[b]Institute of Language and Communication, University of Southern Denmark

Abstract

The Fyntour multilingual weather and sea dialogue system provides pervasive access to weather, wind and water conditions for domestic and international tourists who come to fish for seatrout from the coasts of the Danish island of Funen. Callers access information about high and low waters, wind direction, etc. via spoken dialogues in Danish, English or German. We describe the solutions we have implemented to deal with number format data in more languages. Additionally, we show how the translation of free text data from Danish to English is handled by a newly developed machine translation system, which is rule-based as opposed to most existing, statistically based MT-systems. In conclusion, an assessment – derived from the analysis of recorded calls and user feedback - is made concerning the usability of the system.

1. Introduction

Tourism is one of the major industries on the Danish island of Funen - in particular, a substantial number of international tourists come to fish for seatrout from the coasts of Funen. The main purpose of the Fyntour dialogue system, which is available at phone: 0045 70 22 22 74 at standard call charges, is to provide pervasive access to weather, wind and water conditions for anglers who use information about high and low waters, wind direction, etc. to plan their fishing trips. The general public can be viewed as a secondary user group since easy access to weather forecasts is useful to most people.

The decision to use a spoken dialogue system was based on the following observations:

1. Most tourists do not bring laptops or PDA's and therefore have limited or no access to weather information on the web.
2. Tourists normally have mobile phones, or access to a telephone in their hotel room.

Other current 'voice forecast' systems such as the Weatherdial[1] system of the Irish Meteorological Service and Wendy[2] - beach-specific forecasts for windsurfers in North America, Hawaii, Baja, and the Caribbean – also use text-to-speech (TTS) to deliver information over the phone. In the Weatherdial system, forecasts for the Irish provinces are mapped to separate access numbers, whereas Wendy uses touch-tones (DTMF) to let subscribing users choose specific types of information from locations that are pre-selected at the Wendy website. The public service Fyntour system uses automatic speech recognition (ASR) to let users navigate via a dialogue that combines system

* Corresponding author:
E-mail addresses: ahlmann@voicetech.dk (J. A. Hansen), eckhard.bick@mail.dk (E. Bick)
[1] "Weatherdial". [Online]. Last updated: 24 November 2005. <http://www.met.ie/aboutus/weatherdial/default.asp>: The Irish Meteorological Service.
[2] "Wendy". [Online]. Last updated: 2005. <http://www.iwindsurf.com/services.iws?genID=46>: WeatherFlow Inc.

direction and mixed initiative. Speech-enabled interaction has the following advantages over DTMF input:

1. Users can make more choices with one sentence, e.g. “North and east Funen, please.” selects information from 2 areas.
2. Users can navigate the information system in a more flexible way. Expert users can access the desired information directly, while first-time callers can use the system as a guide.
3. Most people prefer to use the telephone for speaking. Surveys show that users often find it troublesome to press the small keys of mobile phones.

In the MIT Jupiter System, a multilingual, interlingua-based weather information system (Zue et al. 2000), the system’s weather information is mainly extracted from websites such as CNN and The National Weather Service. In contrast, the weather prognoses and water information of the Fyntour system are specifically tailored to the application and the geographical area of Funen. The Danish Meteorological Institute (DMI) produces a general forecast in a text format for the island, but also hour-by-hour weather prognoses and tide tables for the 4 coastal areas shown below:

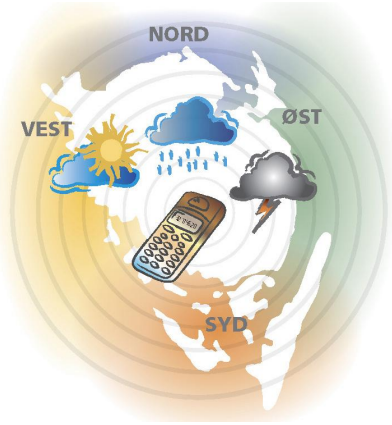

Fig. 1 – Map of Funen

Additionally, local biologists have set up electronic measuring devices in the waters surrounding the island - data showing water temperature and the exact water level is sent to our application server at regular intervals. Although the almost circular geography of Funen is especially well suited to this compass-selection approach, the design should be applicable to almost any tourist area with few modifications. Since the NSEW distinctions are international, they help to bypass the following potentially problematic issues: first, tourists are probably not familiar with local place names and may be unable to relate the position of their residence to these place names. Second, the ASR does not have to deal with ‘foreign’ pronunciations of local place names, e.g. tourists trying to pronounce Danish city names such as ‘Fåborg’ / ‘fə bå’ /.

The Fyntour system provides information in Danish, English and German. A substantial amount of data is received and handled in an interlingua format, i.e. data showing wind speed (in meters per second) and precipitation (in mm) are language-neutral numbers which are simply converted into language-specific pronunciations by specifying the locale of the speech synthesis in the VoiceXML, as an attribute in <vxml> and/or <prompt> tags, e.g.

```
<prompt xml:lang="da-DK"> 1 </prompt>
<prompt xml:lang="en-GB"> 1 </prompt>
<prompt xml:lang="de-DE"> 1 </prompt>
```
→ ”en”
→ ”one”
→ ”ein”

In Germany, wind speed is normally measured using the Beaufort scale (vs. the Danish m/s norm), while visitors from English speaking countries are accustomed to the 12-hour clock (vs. the continental European 24-hour clock). These cultural preferences can be catered for by straightforward conversions of the shared number format data – performed by the application logic that generates the dynamic VXML output of the individual languages.

However, the translation of dynamic data in a free text format, from Danish to English and Danish to German, – such as the above-mentioned 24-hour forecasts, written in Danish by different meteorologists – is more complex. In the Fyntour system, the Danish-English translation problem has been solved by a newly developed machine translation (MT) system. The Constraint Grammar based MT-system, which is rule-based as opposed to most existing, statistically based systems, is introduced in section 3.

2. System Architecture – Portability

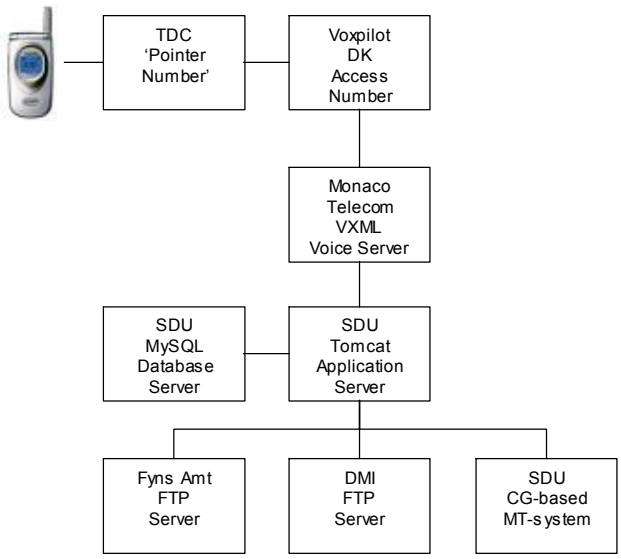

Fig. 2 – Fyntour System Architecture

The highly distributed structure of the Fyntour system does not cause any noticeable delays in the flow of the dialogue, e.g. the system response time for dynamic output produced by the remote application server equals the response time of the static navigational prompts.

While voice technology is a rapidly developing field, both in terms of quality and in terms of pricing (hardware and software), establishing an in-house telephony gateway and voice server with multilingual speech engines can still be quite expensive. Therefore, we have opted for a tentative outsourcing solution, in which Voxpilot[3] hosts

[3] “Voxpilot”. [Online]. Last updated: 9 August 2005. <http://www.voxpilot.com>:

the voice interface on a voice server platform operated by Monaco Telecom. As shown in figure 2, the voice server in Monaco communicates with an application server at the University of Southern Denmark (SDU). The SDU application server returns the user-requested information - in the form of VXML sub-dialogues - to the voice server via the HTTP-protocol. Sub-dialogues are similar to programming language functions or methods with or without a return value. In our case they are called by specifying a remote URL within a <form> element. When the Form interpretation Algorithm (FIA), which decides the order of VXML processing, exits a sub-dialogue, it will automatically transition back to the VXML form from which the sub-dialogue was called. This facilitates the portability of the VXML application – if, at some point, it is decided to use a different voice server platform, it is not necessary to change any URL's in the application server logic.

The access telephone number of the Fyntour weather and sea information system has been published in a wide range of media: newspapers, local TV and radio, fishing guides, etc. It is obviously important that the access number remains constant, while all other application components should be flexible. To this end, we use a 'pointer number', which can be set to point to any other number. As shown in figure 2, the pointer number currently refers calls to a Danish Voxpilot access number, from where calls are transported to Monaco Telecom (without extra call fees or delays).

To further ensure the portability of the system, the backend technologies are open source, such as a Tomcat Java Servlet Container and a MySQL database server running on Linux platforms. Care has been taken to use portable W3C[4] industry standards such as Voice Extensible Markup Language (VoiceXML), Speech Synthesis Markup Language (SSML) and ASR grammars that follow the Speech Recognition Grammar Specification (SRGS). It should be noted that while the SRGS specifies both an ABNF and an XML (.grxml) form, only the XML form is truly portable, e.g. between an IBM and a Nuance voice server platform. Although the Voxpilot Voice Server has been certified for VoiceXML 2.0 conformance by the VoiceXML Forum, semantic interpretation of user utterances must be implemented in the the Nuance Grammar Specification Language (GSL) tag format, which is not portable to e.g. an IBM voice server[5].

3. CG-based MT System

The Danish-English MT module, Dan2eng, is a robust system with a broad-coverage lexicon and grammar, which in principle will translate unrestricted Danish text or transcribed speech without strict limitations to genre, topic or style. However, a small benchmark corpus of weather forecasts was used to tune the system to this domain and to avoid lexical or structural translation gaps, especially concerning time and measure expressions, as well as certain geographical references and names.

Methodologically, the system is rule-based rather than statistical and uses a lexical transfer approach with a strong emphasis on source language (SL) analysis, provided by a pre-existing Constraint Grammar (CG) parser for Danish, DanGram (Bick 2001). Contextual rules are used at 4 levels:

Voxpilot/Eurekasoft.

[4] "W3C". [Online]. Last updated: 1 December 2005. <http://www.w3.org>: The World Wide Web Consortium (W3C).

[5] The IBM WebSphere Voice Server supports the W3C Semantic Interpretation for Speech Recognition (SISR) specification.

- CG rules handling morphological disambiguation and the mapping of syntactic functions for Danish (approximately 6.000 rules)
- Dependency rules establishing syntactic-semantic links between words or multi-word expressions (220 rules)
- Lexical transfer rules selecting translation equivalents depending on grammatical categories, dependencies and other structural context (16.540 rules)
- Generation rules for inflexion, verb chains, compounding etc. (about 700 rules)
- Syntactic movement rules turning Danish word order into English word order and handling subclauses, negations, questions etc. (65 rules)

At all levels, CG rules may be exploited to add or alter grammatical tags that will trigger or facilitate other types of rules.

As an example, let us have a look at the translation spectrum of the weatherwise tedious, but linguistically interesting, Danish verb *at regne (to rain),* which has many other, non-meteorological, meanings *(calculate, consider, expect, convert ...)* as well. Rather than ignoring such ambiguity and build a narrow weather forecast MT system or, on the other hand, strive to make an "AI" module *understand* these meanings in terms of world knowledge, Dan2eng chooses a pragmatic middle ground where grammatical tags and grammatical context are used as *differentiators* for possible translation equivalents, staying close to the (robust) SL analysis. Thus, the translation *rain (a)* is chosen if a daughter/dependent (D) exists with the function of situative/formal subject (@S-SUBJ), while most other meanings ask for a human subject. As a default[6] translation for the latter *calculate (f)* is chosen, but the presence of other dependents (objects or particles) may trigger other translations. *regne med (c-e),* for instance, will mean *include,* if *med* has been identified as an adverb, while the preposition *med* triggers the translations *count on* for human "granddaughter" dependents (GD = <H>), and *expect* otherwise. Note that the *include* translation also could have been conditioned by the presence of an object (D = @ACC), but would then have to be differentiated from (b), *regne for ('consider').*

regne_V[7]
(a)D=(@S-SUBJ) :rain;
(b)D=(<H> @ACC) D=("for" PRP)_nil :consider;
(c)D=("med" PRP)_on GD=(<H>) :count;
(d)D=("med" PRP)_on :expect;
(e)D=(@ACC) D=("med" ADV)_nil :include;
(f) D=(<H> @SUBJ) D?=("på" PRP)_nil :calculate;

It must be stressed that the use of grammatical relations as translation differentiators is very different from a simple memory based approach, where chains of words are matched from parallel corpora. First, the latter approach - at least in its naïve, lexicon-free version - cannot generalize over semantic prototypes (e.g. <H> for human) or syntactic functions, conjuring up the problem of sparse data. Second, simple collocation, or co-occurrence, is much less robust than functional dependency relations that will

[6] The ordering of differentiator-translation pairs is important - defaults, with fewer restrictions, have to come last. For the numerical value of a given translation, 1/rank is used. The example lacks the general, differentiator-less default provided with all real lexicon entries.
[7] The full list of differentiators for this verb contains 13 cases, including several prepositional complements not included here *(regne efter, blandt, fra, om, sammen, ud, fejl ...)*

allow interfering material such as modifiers or sub-clauses, as well as inflexional or lexical variation.

In the confines of this article, no detailed description of the individual translation stages is possible, but figure 3 illustrates the modularity of the system in a schematic way. For a closer look, see http://beta.visl.sdu.dk (demo, documentation, DanGram papers).

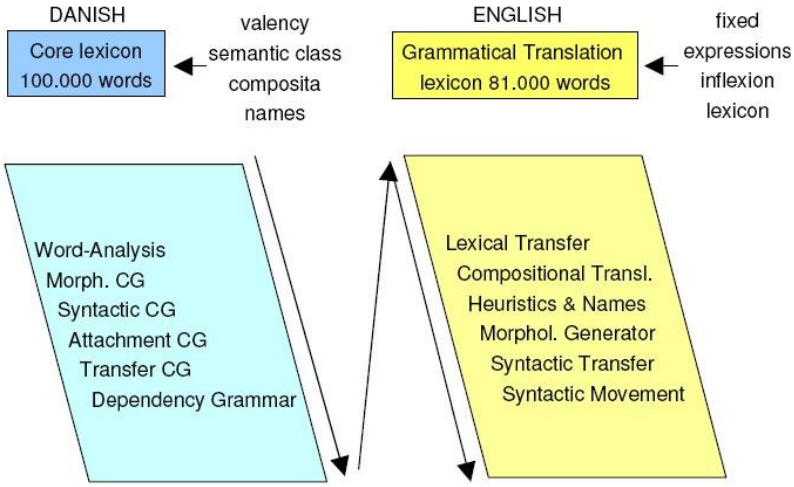

Fig. 3

4. Dialogue Design and Assessment

The prototype dialogue structure of the Fyntour system was a result of close cooperation between tourist and fishing experts, meteorologists, biologists, computational linguists, native speakers of English, German and Danish, and voice technology programmers. Figure 4 indicates the current dialogue structure in which each language is handled separately, following an initial system identification and language selection dialogue. As Delgado and Araki agree: “The most straightforward method for creating a multilingual dialogue system is to prepare the ASR and NLU modules for each language and make them generate a language independent common semantic representation.” (Delgado, R. L. & Araki, M. 2005: 88). E.g. if the ASR modules of all three languages produce a common semantic representation for user-selected compass directions, it greatly facilitates interaction between the VXML interface and the backend logic, which returns dynamic VXML according to user-selected parameters.

Novice callers can interact via a hierarchically organised dialogue structure, which allows users to navigate via phonetically distinct, binary choices. The aim here is to reduce cognitive load and to maximise ASR performance. In the system introduction of the individual languages, users are informed that it is possible to interrupt the system at any time and say the ‘shortcut’ name of the desired information section. This allows the experienced user to skip some navigational turns and therefore to access the desired information more directly.

An alpha-version was made available to a small group of primary users (local anglers), who provided valuable feedback concerning the usability of the system. Additionally, log files from this initial testing provided valuable information about ASR performance, etc. Subsequently, a beta version was made publicly available, when the access number was shown during a TV-documentary on the Fyntour system. 100 Danish calls were logged immediately after the TV presentation and subsequently analysed to further improve the wording of prompts, grammar coverage and to fine tune

confidence levels for acceptance of user utterances. Though these first callers provided a worst case scenario - in the sense that many callers where just curious to talk to a computer, and not seriously interested in high and low waters – most people, surprisingly, did not experiment, but were cooperative and followed the mainly system-directed dialogue nicely.

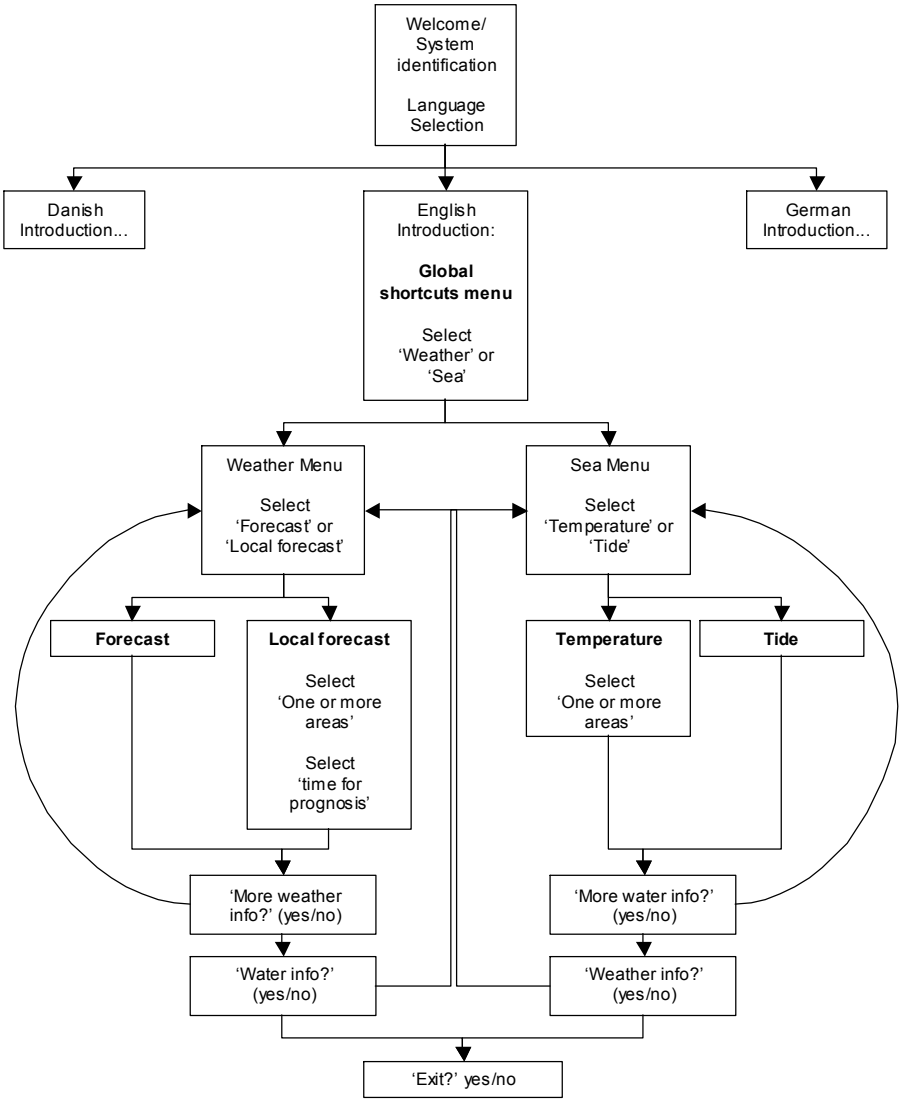

Fig. 4

However, the log files showed that minor additions were needed to the ASR grammars and indicated that certain prompts should be shortened in order to improve clarity and efficiency. To test the effect of these changes, 100 Danish calls[8] were collected and analysed one month later. These data were relatively positive: if a small number of callers who experienced problems due to bad mobile phone connections is discounted, callers had a high rate of task completion (95-98 %) with an ensuing low abandonment rate, i.e. the rate of callers who hang up without getting the desired information. The last

[8] At the time of writing, the Fyntour access number had not been publicly announced to English and German speaking tourists.

set of recorded dialogues contains more cases of multiple calls from the same person. Typically, these callers have learned to use barge in and to apply the shortcuts of the dialogue.

The communicative competence of the system is quite limited and the dialogue structure obviously does not have the flexibility of human-to-human conversation. However, log data show a high rate of task completion, while user comments on the system in a fishing web discussion forum indicate a relatively high degree of satisfaction and even enthusiasm about the dialogue system. These comments are probably influenced by the fact that dialogue systems are relatively new and rare in Denmark and the fact that the system provides easy access to specific information that is not accessible elsewhere. Another aspect may be that people often find themselves in a variety of situations in which conversation or communication is restricted in ways that are similar to human-computer dialogues, e.g. classroom teaching, job interviews, doctor consultations, and parents talking to small children. The latter form of communication has been labelled 'parentease', and is characterised by such features as higher pitch, exaggerated intonation, shorter utterances and simpler syntax. This may be one explanation why our data show that people quickly adopt a 'computerease' style of speech in which they pay more attention to what they say and speak with a clearer pronunciation.

6. References

1. Bick, Eckhard. 2001. "En Constraint Grammar Parser for Dansk". In: Peter Widell & Mette Kunøe (eds.), *8. Møde om Udforskningen af Dansk Sprog, 12.-13. oktober 2000*, pp. 40-50, Århus University.
2. Delgado, R. L. & Araki, M. 2005. *Spoken, Mulitilingual and Muiltimodal Dialogue Systems – Development and Assessment*. Chichester: John Wiley & Sons, Ltd.
3. Ellis, A. & Beattie, G. 1986. *The Psychology of Language and Communication*. London: Lawrence Erlbaum Associates.
4. Larson, James A. et al. 2005. "Ten Criteria for Measuring Effective Voice User Interfaces". [Online]. *Speech Technology Magazine, November/December 2005, Vol. 10, No. 6*. <http://speechtechmag.com/issues/>.
5. Sharma, C. & Kunins, J. 2002. *VoiceXML – Strategies and Techniques for Effective Voice Application Development with VoiceXML 2.0*. New York: John Wiley & Sons, Inc.
6. Suchman, Lucy A. 1987. *Plans and Situated Actions – The Problem of Human Machine Communication*. Cambridge: CUP.
7. Turing, Alan M. 1950. "Computing Machinery and Intelligence". *Mind*, Vol. 59, No. 236, pp.433-460.
8. "Voxpilot". [Online]. Last updated: 9 August 2005. <http://www.voxpilot.com>: Voxpilot/Eurekasoft.
9. "Weatherdial". [Online]. Last updated: 24 November 2005. <http://www.met.ie/aboutus/weatherdial/default.asp>: The Irish Meteorological Service.
10. "Wendy". [Online]. Last updated: 2005. <http://www.iwindsurf.com/services.iws?genID=46>: WeatherFlow Inc.
11. "W3C". [Online]. Last updated: 1 December 2005. <http://www.w3.org>: The World Wide Web Consortium (W3C).
12. Zue, V. et al. 2000. "Jupiter: A Telephone-Based Conversational Interface for Weather Information". *Trans. Speech and Audio Proc.*, 8(1), 85-96, (Institute of Electrical and Electronic Engineers).