Proceedings of ICSLP 2002, Denver, September 16–20, 2002.
Paper 1557, Vol. 2, pp. 837–840.
CD-ROM ISBN 1-876346-40-X

INGRESSIVE SPEECH AS AN INDICATION THAT HUMANS ARE TALKING TO HUMANS (AND NOT TO MACHINES)

Robert Eklund †‡

† Telia Research AB, Farsta, Sweden
‡ NLPLab, Department of Computer Science, Linköping University, Sweden
Robert.H.Eklund@telia.se

ABSTRACT

Pulmonic ingressive speech is often mentioned anecdotally in the linguistic research. Most previous studies investigating the phenomenon have stressed the paralinguistic function of ingressive speech (IS). This paper studies IS in two corpora of spontaneous Swedish speech. Eight subjects made business travel bookings in two data collections. In one corpus the subjects talked with a real, human travel agent; in the other they spoke with what they believed was a computer, played by a professional actor. The results show that all subjects made use of IS in the human–human setting, while no one used IS in the human–machine setting. These results strengthen the notion that IS is a speech phenomenon that is truly associated with human interactions. The results are discussed from the perspective of possible underlying factors, including discourse structure, gender issues, and possible enhancements in automatic speech-based dialog systems.

1. INTRODUCTION

This paper deals with the production of pulmonic ingressive speech (as opposed to glottal ingressive or velaric ingressive speech). Despite ubiquitous anecdotal information, not much seems to have been written on the subject, although it is possible to find reports of ingressive speech (IS) in e.g., all North European countries.[1] Most of these works have discussed the issue primarily from a discourse angle, and stressed the paralinguistic function IS serves, in particular its interactional role in spontaneous human–human (HH) conversation, e.g., as a specific feedback marker, an extra strong confirmation, a topic-closing marker, or as a "bonding" device. It is also often claimed that women make use of IS more often than men.

This paper will also focus on the interactional aspects of IS. While previous studies have examined human–human conversations, the present study compares human–human dialogs with human–machine dialogs over a telephone line, in which subjects performed identical tasks in all other respects.

Reeves & Nass [9] claim that humans interact with media in basically the same way they interact with other human beings. Based on the observations in this study, potential implications for this claim are discussed, and possible enhancements in automatic speech-based dialog systems are suggested.

[1] An exhaustive review of the literature on ingressive speech will be given in Robert Eklund: *Ingressive Speech: What, How, Where, Who, and Why* (in preparation).

2. PREVIOUS STUDIES

It is often said that IS is typical of **Scandinavian** speakers, and that non-Scandinavians who encounter IS in Swedish or Norwegian get the impression that the speaker either is about to suffocate or is suffering from severe shock [3:33], is surprised, or has a heart condition [5]. Allwood reports IS in corpora of spontaneous **Swedish** speech and gives examples of the IS particles "ja", "jaha", "jo" (variants of 'yes') and "nä" ('no') [2:97]. Another study that mentions ingressive feedback words in Swedish is Landqvist [6]. Stølen [11] describes the use of ingressive "ja" (yes) in **Danish**, while Hakulinen [4] analyzes IS in **Finnish**. Pitschmann [8] discusses the phenomenon as it occurs in **German** and **Scandinavian** languages, including **Icelandic**, and Peters [7] describes the paralinguistic role of IS in **Maine English** and **Norwegian**. In a recent study, Kobayashi [5] has made an extensive analysis of IS in Norwegian from a discourse perspective, while Shorrocks [10] analyzes IS in **Newfoundland English**, and also reports IS in **British** dialects and **Irish.**

3. DATA AND METHOD

Two corpora of spontaneous Swedish telephone speech were used. Both were collected in 1997.

In the first corpus, **WOZ2**, forty-six subjects booked business travels, speaking with what they believed was an automatic travel booking system. In fact, an actor (or 'wizard') pretended to be the speech synthesizer/computer, reading out small set of scripted utterances. This means that the quality of the "speech synthesis" was perfect, containing no acoustic artifacts, typical of state-of-the art speech synthesizers. This corpus will be referred to as the Human–Machine (HM) corpus.

In the second corpus, **Nymans**, eight subjects—all of whom had participated in WOZ2—performed the same exact tasks as in WOZ2, this time speaking with real, human travel agents at the travel agency Nyman & Schultz, in Haninge, Sweden. This corpus will be referred to as the Human–Human (HH) corpus. Since WOZ2 was conducted six months prior to Nymans, it can be assumed that no learning effects show up in the data. Summary statistics for both corpora are presented in Table 1.

The subjects/agents were sitting in a room with a headset phone (used for DAT recordings), as well as a normal, landline telephone (used to tap the telephone line). Thus, the setting was naturalistic for the agents, near-natural for the subjects.

Proceedings of ICSLP 2002, Denver, September 16–20, 2002.
Paper 1557, Vol. 2, pp. 837–840.
CD-ROM ISBN 1-876346-40-X

Table 1: Summary statistics for WOZ2 and Nymans. WOZ2 figures are given separately for the eight subjects who participated in Nymans. (Thus the "WOZ2/also in Nymans" column is consequently properly included in the WOZ2 data.)

	WOZ2	WOZ2 also in Nymans	Nymans
No. subjects	46 (32/14F)	8 (6M/2F)	
No. dialogs	140	24	24
No. utterances	3444	625	1730
No. utts./dial.	25	26	72

The subjects were given the tasks in mainly iconic form, using a variety of symbols to denote hotels, prices, trains etc. This was done to avoid linguistic biasing, which is a known effect when written instructions are used. All subjects performed three or four tasks each.

All data were labeled and analyzed by the author, using waves™ on a Sun work station.

4. RESULTS

4.1. Frequency distribution of ingressive speech items

All instances of IS are listed in Table 2.

Table 2: Ingressive speech for clients in WOZ2 (HM) and for clients and agents in Nymans (HH). The question mark indicates that the item is barely audible, even with headphones.

Client	Gender		HM	HH	
	Client	Agent		Client	Agent
1 (LL)	M	F	0	1	0
2 (DM)	M	F	0	2	0
3 (MN)	M	F	0	4	0
4 (FS)	F	F	0	34	12
5 (HS)	M	M	0	12	3
6 (BU)	F	M	0	1	0
7 (RO)	M	M	0	6	0
8 (MS)	M	M	1?	8	0
		Σ	1?	68	15

As can be seen, while all subjects produced IS in Nymans, no one, with one possible exception, made use of ingressives in WOZ2. The sole example of IS in HM is barely audible, even with headphones (DAT recording), and is probably similar to something Stølen calls "inner-directed" [11:672].

As a control, the frequency of IS was checked in the entire WOZ2 corpus—i.e., the other 38 subjects (26M/12F), 116 dialogs and 2819 utterances—without finding a single instance of IS.

4.2. Typology of ingressive speech items

All instances of IS are typical feedback words, i.e. ingressive variants of (normally egressively produced) words such as "ja" ('yes') or "nej" ('no')—which occur both in voiced and unvoiced form—affirmative "mm" (always voiced) or other feedback words like "bra" ('good', 'ok'). The distribution of types is shown in Table 3.

The data in Table 3 confirm the literature in that "ja" is by far the most common ingressive feedback signal in Swedish, and other Scandinavian languages, followed by the likewise (affirmative) feedback "mm". Also noteworthy is that of all "ja" produced, over 10% are ingressive, making it the 33rd most common word (tokens) in the corpus, ranking alongside egressive "nej" ('no'), "med" ('with') and "inte" ('not')—all of which having 51 tokens.

Table 3: Relative frequency of ingressive and corresponding egressive feedback markers pooled for clients and agents in WOZ2 and Nymans. Note that both clients and agents also make use of a variety of other, egressive-only, feedback words.

Type	Client		Agent	
	Ingressive	Egressive	Ingressive	Egressive
ja ('yes')	51	658	14	27
mm ('mm-hm')	12	361	–	–
a ('yeah')	1	3	–	–
nej ('no')	1	52	1	3
joråå ('oh yes')	1	1	–	–
jadå ('indeed')	1	3	–	–
bra ('good')	1	69	–	–
Σ	68	1162	15	30

4.3. Human–Machine vs. Human–Human

The highly scripted system utterances in WOZ2 included none of the words in Table 2, or indeed any other similar feedback words, and one must ask whether feedback signaling is needed from the agent to elicit feedback signaling in the subjects. Do subjects/clients give feedback even if the system does not make use of similar linguistic means? The occurrence of client-produced feedback words broken down for WOZ2 and Nymans is shown in Table 4.

Table 4: Frequency of feedback words in WOZ2 (HM) and Nymans (HH). As "ja" (yes) counts all variants, such as "jaha".

Client	"ja", "jo" etc.		"mm"		"nej"	
	HM	HH	HM	HH	HM	HH
1 (LL)	4	55	10	28	5	8
2 (DM)	23	69	0	81	9	3
3 (MN)	17	16	9	33	9	3
4 (FS)	11	109	16	121	4	14
5 (HS)	1	112	0	44	2	5
6 (BU)	2	96	0	24	2	7
7 (RO)	1	116	0	34	4	6
8 (MS)	15	143	0	0	4	7
Σ	74	716	35	377	39	52

Although there are significant differences between the HM and HH settings, the interesting observation, however, is that all subjects make use of feedback words in the *both* the HH and the HM corpora. This implies that feedback from the system is *not* a prerequisite for feedback in the clients. However, it must be noted that not all these words, or their particular function at given places, have identical "status". While some instances of e.g., "ja" are clear cases of feedback signals, some are simply replies to yes/no-questions, with little paralinguistic function.

Proceedings of ICSLP 2002, Denver, September 16–20, 2002.
Paper 1557, Vol. 2, pp. 837–840.
CD-ROM ISBN 1-876346-40-X

4.4. Distribution of IS in the dialogs

The distribution of IS in the dialogs are shown in Table 5

Table 5: Distribution of ingressive speech in Nymans (HH) broken down by task/dialog for both subjects/clients and agents.

Client	Dialog number	IS	
		Client	Agent
1 (LL)	1	0	0
	2	0	0
	3	1	0
2 (DM)	1	0	0
	2	1	0
	3	1	0
3 (MN)	1	1	0
	2	2	0
	3	1	0
4 (FS)	1	6	1
	2	16	1
	3	12	10
5 (HS)	1	3	0
	2	3	3
	3	6	0
6 (BU)	1	0	0
	2	0	0
	3	1	0
7 (RO)	1	3	0
	2	2	0
	3	1	0
8 (MS)	1	2	0
	2	5	0
	3	1	0

In all four dialogs where both client and agent use IS, it is the client who initiates its use. Client 4 makes very frequent use of IS, and it is notable that it is only in those dialogs the (female) agent makes use of IS. However, the fact that the client makes use of IS is not enough to elicit the same behavior in the agents, since client 5 makes frequent use of IS in the third dialog, without eliciting IS in the (male) agent

5. DISCUSSION

The most striking observations that can be made from the data presented above are that 1) all subjects employ IS in the HH dialogs, and 2) no subjects employ IS in the HM dialogs This lends strong support to the notion that something "human" is lacking in the interaction with the (assumed) machine. Some potential factors will be discussed below.

5.1. The role of system feedback

Feedback is typical of human interaction, and Stølen points out that ingressive feedback is typical of telephone speech, where visual feedback is missing [11:671]. (However, Hakulinen [4:52] did not find IS in her telephone data.) As was shown, the fact that the WOZ2 system provided no feedback signals is surely to a large degree responsible for the lack of IS in WOZ2, but cannot be the only explanation, since the subjects still produce feedback signals in WOZ2, only fewer.

Another function of feedback markers, and a possible characteristic trait of IS, is that they serve some kind of "bonding" function [11:674]. This could also explain their absence in WOZ2: humans simply don't bond with machines!

5.2. The role of dialog structure

While the wizard in WOZ2 employed a small set of highly scripted utterances, the dialogs in Nymans were fully free, and not only were more utterances used to solve the same tasks (cf. Table 1), the fact that clients could speak more freely without being misunderstood or in any way restricted surely allowed for more genuine speech. Thus, it could be the case that a system that simply allows for freer, less constrained, conversation could elicit IS in the clients.

5.3. The role of gender

In the literature, it is often claimed that women employ more IS [4:52, footnote 6][11:671][5:95]. The data in the present study are too scarce to either corroborate or refute any assumptions about gender differences. The clearest observation, however, is that *all* subjects use IS in the HH corpus. However, the two dialogs where the highest frequency of IS are found are both *same-gender* dialogs, confirming Stølen [11:674], who points out that discourse particles in general rise in frequency in same-gender conversation. Also, while the three instances of the (male) agent's IS in dialog 5 are produced in a row (at 806, 860 and 906 seconds into the dialog), without any intervening IS from the client, the (female–female) dialog 4 includes five instances where the client and the agent produce IS in an intertwined way within less than three-second time frames, lending support to Stølen's conclusion that women use IS particles to regulate their conversation. [11:676].

5.4. The role of linguistic function

Several studies have discussed the function of IS particles. Hakulinen [4] discusses the functions *affective, self-directed, turn-taking, responsive* and *topic-closing* in her data on Finnish. Stølen, analyzing Danish, mentions the functions *affirmative, inner-directed*, and *aligning* [11].

The main function of IS in Swedish is feedback signaling. Allwood lists IS as one of several possible "phonological and morphological operations" used on feedback words in Swedish [1:19]. Landqvist [6:142–146], also studying Swedish, argues that the main functions of IS are *argumentative*, restating information already mentioned, and *topic-closing*, signaling that everything that could be said within a discourse fragment is already said.

Kobayashi [5:27–29] lists seven possible functions of IS in Norwegian: 1. *Feedback*. 2. *Self-reflecting*. 3. *Irritation* or *impatience*. 4. *Friendliness* or *intimacy* (cf. Hakulinen [6:52]). 5. *Confirmative* (cf. Landqvist's argumentative). 6. *Concluding* (cf. Landqvist's topic-closing). 7. *Turn-taking*. Kobayashi [5:95] finds support for functions 1, 6, 5 and 7, in that order of (falling) frequency, which more or less corroborates Landqvist's findings for Swedish.

The results of a functional analysis of the data in this paper are shown in Table 6.

The present data seem to confirm some of the findings in the literature. The **feedback** function is by far the most common. It must, however, be pointed out here that words like "yes" are also used as simple replies to yes/no-questions, with very little obvious paralinguistic function, and that all such instances have been counted as feedback here. The **closing** function is the second most frequent function encountered in the present data,

Proceedings of ICSLP 2002, Denver, September 16–20, 2002.
Paper 1557, Vol. 2, pp. 837–840.
CD-ROM ISBN 1-876346-40-X

which agrees with Kobayashi's results. Kobayashi, however, found no clear instances where IS was used to signal **intimacy**. Although interpretation of function is difficult, the five instances of "aligned" IS in dialog 4 could be said to serve the function of intimacy. The **confirmation** function is also clearly found in most dialogs. The other functions listed by Kobayashi are not found in the present data.

Table 6: Functions of IS in the present corpora. The data for subjects 1–3 and 6–8 are pooled, while the figures for dialog 4 and 5 are broken down for clients and agents. Note that the some figures exceed the sum totals in Table 2, since instances of IS may serve more than one function.

Function	1–3/6–8	4		5	
	Clients	Client	Agent	Client	Agent
Feedback	16	21	7	7	–
Closing	5	7	3	4	–
Intimacy	–	5	5	1	–
Confirming	1	4	3	–	3

6. CONCLUSIONS

Reeves & Nass [9] claim that people tend to treat and react to media (TV, computers etc.) much the same way they treat human beings, based on the assumption that human behavior is profoundly social for evolutionary reasons. One example that confirms this in the present study is the observation that paralinguistic feedback does occur in WOZ2. However, other observations in this study seem to run counter to a strong interpretation of Reeves & Nass, since the results seem to suggest that human beings indeed *do* make a difference with regard to how they treat media, making use of ingressive feedback only in the HH setting. This is even more interesting given that feedback signals like "yeah" to a large degree are subconscious. It must, however, be borne in mind that the more constrained dialog in WOZ2 surely is partly responsible for these results. This of course raises the question whether one could make computers more "human" or "natural" by having them use IS. If a computer were to use ingressive feedback (or other possible functions of IS) signals—naturally at the "right" places— would that elicit IS from humans? If it would, then IS could be used as a metric to judge the "naturalness" of an automatic system. If a computerized system is capable of eliciting IS in humans, then one could assume that the users feel at ease with the system. In any case, in order to enhance the naturalness of automatic systems, they should be able to make use of IS, at least in those languages where IS is a normal feedback signal.

Not only could IS play a role in the perceived naturalness in *speaking* computers. If an automatic system would elicit IS in the user, then IS would also have to be considered in the training of automatic speech recognizers. As has been shown, some speakers make very frequent use of IS. Given how common feedback markers are, care should be taken to include them in the training material for recognizers. To that end, it is important to remember that IS will likely not occur in any other settings than genuine, spontaneous, human conversation—which is probably one reason why it is not often noted in the literature. Thus, IS cannot be elicited in most kinds of formal settings, typical for how training data are recorded for speech recognizers. Moreover, not only does IS differ from egressive speech acoustically, it also serves specific linguistic *functions*, and consequently cannot be discarded without good reason. It would be extremely interesting to see whether IS in the machine would elicit IS in the subjects. If the machine were to employ such a truly human trait, would humans be lured into "bonding with machines"?

7. ACKNOWLEDGEMENTS

Thanks to Martin Eineborg, Joakim Gustafson, Anders Lindström and Åsa Wengelin for comments on earlier drafts of this paper. Thanks to Michael Kiefte for proofing.

8. REFERENCES

[1] Allwood, Jens. 1988. The Structure of Dialogue. In: M. M. Taylor, F. Néel & D. G. Bouwhuis (eds.): *The Structure of Multimodal Dialog II*, Amsterdam, John Benjamins, 3–24.

[2] Allwood, Jens. 1988. Om det svenska systemet för språklig återkoppling. In: Per Linell, Viveka Adelswärd, Torbjörn Nilsson & Per A. Pettersson (eds.): *Svenskans Beskrivning* 16, Vol. 1, Linköping University, 89–106.

[3] Allwood, Jens. 1982. Finns det svenska kommunikationsmönster? In: *Vad är svensk kultur? Papers in Anthropological Linguistics* 9, Department of Linguistics, Gothenburg University, 6–49.

[4] Hakulinen, Auli. 1993. Inandningen som kulturellt interaktionsfenomen. In: A.-M. Ivars et al. (eds.): *Språk och social kontext*. Meddelanden från institutionen för nordiska språk och litteratur vid Helsingfors Universitet, serie B:15. Helsinki University, 49–67.

[5] Kobayashi, Nazuki. 2001. *Ingressivt "Ja". Ja på innpust – ikke tegn på overraskelse eller dårlig hjerte*. MA thesis, Institutt for Lingvistikk og Litteraturvitenskap, Bergen University.

[6] Landqvist, Håkan. 2001. *Råd och ruelse. Moral och samtalsstrategier i Giftinformationscentralens telefonrådgivning*. PhD thesis, Department of Scandinavian Languages (Institutionen för Nordiska Språk), Uppsala University.

[7] Peters, Francis Joseph. 1981. *The Paralinguistic Sympathetic Ingressive Affirmative in English and the Scandinavian Languages*. Unpublished PhD thesis, New York University.

[8] Pitschmann, Louis A. 1987. The Linguistic Use of the Ingressive Air-Stream in German and the Scandinavian Languages. *General Linguistics*, Vol. 27, No. 3, 153–161.

[9] Reeves, Byron & Clifford Nass. 1996. *The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge, Cambridge University Press.

[10] Shorrocks, Graham. Forthcoming. Pulmonic Ingressive Speech in Newfoundland English: A Case of Irish-English Influence? In: Hildegard L. C. Tristram (ed.): *The Celtic Englishes III*. Anglistische Forschungen, Heidelberg Universitätsverlag C. Winter.

[11] Stølen, Marianne. 1994. Gender-related use of the ingressive *Ja* in informal conversation among native speakers of Danish. In: Mary Bucholtz, A. C. Liang, Laurel A. Sutton & Caitlin Hines (eds.): *Cultural Performances: Proceedings of the Third Berkeley Women and Language Conference*, Berkeley Women and Language Group, Berkeley, 668–677.