Sprachtheorie und germanistische Linguistik (www.sugl.eu), 25.1 (2015), 41-74

Ágnes Abuczki

A multimodal discourse-pragmatic analysis of *ugye* (~'is that so?')*

Abstract

This paper presents a corpus-based study of the diachronic development and the multimodal properties of a Hungarian lexical item, *ugye* (~'is that so?') that can be analysed as a multifunctional discourse marker (henceforth DM). DMs are commonly defined as "sequentially dependent elements which bracket units of talk" (Schiffrin 1987: 31), or metalinguistic items that provide information about the segmentation and operation of a discourse (Fraser 1999). It is argued in this work that a multimodal approach is indispensable in communication modelling in order to disambiguate the actual meaning or function of polysemous communicative signals such as DMs. The findings of the corpus queries suggest that the machine-detectable defining features of the analysed DM are its position, the simultaneous performance or cessation of manual gesticulation and the gaze direction of the speaker. These observations have finally led to the development of a decision tree that can distinguish between its two salient functions (checking information and marking explanation), and may later be implemented as an algorithm.
Keywords: discourse markers, pragmatics

* This research was supported by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP-4.2.4.A/ 2-11/1-2012-0001 'National Excellence Program'.

The publication was supported by the SROP-4.2.2.B-15/1/KONV-2015-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund.

1 Objectives, material and methodology

The goal of the paper is threefold: firstly, to describe the historical development of *ugye* (~'is that so?') into a DM; secondly, to uncover its roles in the management of interaction from a discourse-pragmatic perspective[1]; thirdly, to identify verbal, sequential as well as machine-detectable nonverbal features that typically characterize and best distinguish these functions.

The analysis of the diachronic development of the meanings of this lexical item is performed on various historic language resources: primary sources, such as the Historical Corpus of Hungarian (Magyar Történeti Korpusz); and secondary sources, such as etymologic dictionaries (Benkő 1967-1984), historical dictionaries (Ballagi 1872, Szarvas & Simonyi 1893, Szabó 1975-2005), general contemporary dictionaries (Ittzés 2006, Pusztai et al. 2003) and monographs on historical linguistics (Benkő 1992). The synchronic description of the current functional spectrum of *ugye* (~'is that so?') is examined in the Hungarian HuComTech multimodal corpus.

Multiple layer corpus queries and their statistical analyses in SPSS 19.0 address the sequential, suprasegmental and nonverbal properties of the tokens of the selected lexical item. The features in question regard its contextual environment (lexical co-occurrences, presence or absence of surrounding silence), position in the utterance, prosodic features (fundamental frequency, pitch movement) nonverbal-visual markers (facial expression, gaze direction and the presence or absence of accompanying hand movements) and stylistic properties (frequency of use in different discourse genres, in informal conversations and political interviews).

The tools used for corpus analysis include: (1) the annotation software ELAN 4.5.1 (henceforth ELAN) (Brugman & Russel 2004: 2065-2068) for segmenting, tagging and querying DMs, (2) the Praat software for acoustic analysis, (3) the Simple Concordance Program for concordance searches, and (4) SPSS for the statistical analysis of the query results. In short, my methodology combines quantitative and qualitative methods in the multimodal analysis of both audio and video contents of the recordings as well as descriptive and inferential statistical tests during the interpretation phase of the questionnaire and corpus query results.

---

[1] On the discourse-pragmatic view of DMs in English cf. Lewis (2006).

2 The study of *ugye* (~'is that so?') as a DM

First of all, one might ask if *ugye* can be considered as a DM at all. I argue that *ugye* has DM uses since it has relatively free position, it is multifunctional and it encodes procedural meaning (consider its role as a marker of evidentiality on the one hand or marker of new information on the other).

DMs are commonly defined as "sequentially dependent elements which bracket units of talk" (Schiffrin 1987: 31), or metalinguistic items that provide information about the segmentation and operation of a discourse (Fraser 1999). In spite of the fact that Fraser (1999, 2009) identifies them as a subclass of pragmatic markers, I prefer using the term discourse marker because that category (as defined by Schiffrin 1987) is more inclusive as it includes longer, multi-word expressions and nonverbal cues (such as interjections or discourse structuring gestures) as well that also play a crucial role in the successful management of conversation. Schiffrin (1987) describes the role of DMs as "providing contextual coordinates for ongoing talk" that indicate for the hearer how an utterance is to be interpreted. Therefore, DMs are multifunctional pragmatic elements (of heterogeneous word classes) expressing various metacommunicative and cognitive functions. Successful communication requires the ability to infer the intended meaning of ambiguous communicative signals such as multifunctional DMs. It is argued in this work that a multimodal approach is indispensable in communication modelling in order to disambiguate the actual meaning or function of polysemous communicative signals such as DMs.

Concerning its membership in the functional category of DMs, *ugye* displays the following features of the category of DMs:[2]

- non-conceptual meaning, non-propositionality
- procedural meaning, non-compositionality
- context-dependence
- optionality
- weak clause association
- variable scope
- evidentiality ensures that it contributes to argument- and narrative structure as well as inferencing in general

[2] Membership criteria (collected by Furkó 2007) were checked on the relevant material of the HuComTech corpus.

- phonological reduction when used as an evidentiality marker, typically rising intonation when used as a tag question expecting positive reply
- relatively high frequency in speech
- multifunctionality (as it simultaneously marks explanation, evidential information on the one hand, and questions and preference for agreement on the other hand, depending on its position and the illocution of its host utterance)

3 On applying grammaticalization theories and diachronic analysis to account for the emergence of discourse-pragmatic functions

As will be seen in the following examples (1-19) *ugye* used to serve and serves even today multiple functions. The question is how to account for the shifts between meanings and its functional variations. This challenge is addressed in the following section of the paper (sections 3 and 4).

In what follows I will briefly sum up a few leading DM researchers' views on the multifunctional nature of DMs. Three major approaches can be found in the DM literature to explain their variable meanings: (1) the homonymy (also-called maximalist) approach; (2) the monosemy (also-called minimalist) approach; (3) the polysemy approach.

The homonymy or maximalist view can be placed at one end of the theoretical spectrum which is the least popular and least accepted view among the three. The homonymy approach argues that if a given form has a number of seemingly different uses, these various senses represent separate individual lexical items without any relationship between their interpretations.

The monosemy or minimalist approach (Fraser 2006, Thanh 2006) can be placed at the opposite end of the theoretical spectrum. Meaning minimalism attempts to identify an invariant, schematic core meaning of an item from which all uses of a given item can be derived. Variations in the function and use of a given DM arise from its interaction with the context in which it appears.

Proponents of the polysemy view assume that the various interpretations of a single form must be related in a way. Hansen (2006) argues as follows:

> … items which in at least some contexts fulfil a discourse marking function can have more than one meaning on the semantic level, but that these meanings may be related in a motivated – if not necessarily fully predictable – way, such that we may describe as many as possible of the functionally distinct examples of a given homophone/homograph as instantiations of a single, polysemous, lexical item (Hansen 2006: 29).

The difference among the different variants of the polysemy models lies in the fact how different frameworks account for the related nature of the interpretations.

A modified version of the polysemy view (Hansen 2006) that provides the theoretical background in this paper is the core/periphery approach. Taking this approach, Bell (1998: 515-541) counters the assumption that discourse markers are "polysemous". He acknowledges that "markers may be multifunctional" but insists that they have a core function or core pragmatic instruction, and considers those instantiations outside the core to be peripheral. Bell argues that:

> […] the instruction may require a series of inferences derived from the discourse context to arrive at a specific interpretation […] the interpretation of a marker in any one instantiation results from an interaction between its core instruction, the semantic, syntactic and phonological properties of the individual marker, and the context in which the marker appears (Bell 1998: 515-519).

Generally, studies grounded in the framework of the core/periphery approach rely on diachronic analysis in order to identify the core function of the DM and the circumstances that allowed the discourse-pragmatic function emerge. Similarly, I will also use diachronic material when outlining the historical development of *ugye* (~'is that so?') in my case study (sections 4.2 and 4.3).

Grammaticalization theories (Traugott 1995) explain the multifunctionality and the semantic change of DMs based on diachronic research. In Traugott's view (1995), the origin and development of DMs[3] has led to explain the multifunctionality of DMs as a result of grammaticalization. Hopper & Traugott (2003) defines grammaticalization as the systematic change whereby lexical items and constructions develop to serve grammatical functions. Traugott & Dasher (2002) draw the following cline of development towards DM category membership:

---

[3] Traugott (1995) uses the term *pragmatic markers* to refer to the same class of pragmalinguistic elements as I call DMs.

truth conditional meaning → non-truth conditional meaning; content → content/procedural meaning → procedural meaning; non-subjective meaning → subjective meannig → intersubjective meaning; scope within proposition → scope over proposition → scope over discourse

Another term used to describe the semantic change of lexical elements with conceptual meaning from heterogeneous word classes into DMs in pragmaticalization (Erman & Kotsinas 1993, Aijmer 2002). Pragmaticalization involves gradual semantic bleaching (i.e. the loss of semantic meaning) and simultaneous pragmatic enrichment (i.e. gaining pragmatic functions). Among the various branches of pragmatics, historical pragmatics is concerned with the diachronic analysis of the evolution of lexical items with propositional meaning into DMs (items used metacommunicatively).

In my attempt to map the historical development of the propositional lexical item *ugye* into a DM, I will also rely on both diachronic and synchronic data as well as dictionary entries (that is, both primary and secondary sources), and will consequently try to define their current state in the pragmaticalization process using Traugott & Dasher's (2002) model of development.

4 Earlier accounts of *ugye* (~'is that so?'): word class, functional category and etymology

4.1 Uses of ugye today

As far as the word class of *ugye* (~'is that so?') is concerned, it is classified into various part-of-speech categories in different dictionaries, such as adverb, adverbial, modifier,[4] sentence word, question word and (rogative) particle.

Based on the entry in Pusztai et al. (2003: 1391), (1) as an interrogative adverb (kérdő határozószó) it marks the expectation of cooperation and agreement as well as emphasizes this expectation; (2) as a modifier it either (2a) expresses our expectation to receive a positive answer or (2b) expresses politeness / makes a question sound more polite (e.g. "Ön ~[5] ideges?"[6]); on the other hand, (3) it is used as a filler

[4] For a detailed definiton of the category of modifiers, see Kugler 2000.

[5] Following Pusztai et al.'s (2003) symbols, symbol ~ stands for *ugye* in the present section (4.1).

[6] Are you ~ nervous?

without any specific meaning (e.g. "Hát ~ megteszem, de …"[7]). Even the examples and contexts provided by Pusztai et al. (2003) do not help us distinguish the two overlapping senses of *ugye*, (1) "~ jó?"[8] or "~ igazam volt?"[9] as interrogative adverbs and (2) "~igaz?"[10], "~úgy van?"[11] or "Éhes vagy ~?"[12] as modifiers. These examples sound completely synonymous and interchangeable to me. Besides these senses listed in Pusztai et al. (2003), in the course of my synchronic corpus analysis (presented in section 5) I will point out the significance of *ugye* (~'is that so?') in marking evidentiality, a function that is completely missing from dictionary entries. Moreover, its role in connection with narrative story structure as well as lists identified in the empirical part of my study is also absent from reference books.

As far as the synchronic state of *ugye* is concerned, Schirm (2009) highlights that *ugye* retains its original meaning even today to expect the confirmation of the validity of a statement or simply emphasize its validity.

Furthermore, Gyuris (2008) provides an in-depth formal description of *ugye* (referred to as a discourse particle). She argues that its position is relatively free and its presence is not essential for well-formedness. Moreover, it is described with examples that *ugye* can be used in both declaratives and polar interrogatives. As far as its interpretation in interrogatives is concerned, it is a widely accepted view that it produces a biased question and marks the polar interrogative sentence-type by morphological means (H. Molnár 1959, Kiefer 1988, Varga 2002, Keszler 2000). Gyuris (2008) argues that in spite of general consensus, a sentence including *ugye* traditionally described as interrogative is not to be considered an interrogative sentence as far as its form is concerned, but a declarative one, which acquires a question interpretation indirectly, through non-standard, non-assertive intonation with higher pitch.

Concerning its interpretation in declaratives, Péteri (2002, cited in Gyuris 2008) argues that it signals the content of the utterance as

---

7 Well, ~ I'll do it, but…

8 ~ is that good?

9 ~ am I right?

10 ~ is it true?

11 ~ is that so?

12 Are you hungry ~?

common ground among the speakers; accordingly, it is often used in reminders. The etymology of *ugye* described in the following section (4.2) might be accounted for the development of these functions.

4.2 The historical development of the meanings of ugye (~'is that so?')

Concerning the etymology of the lexical item *ugye* (Benkő 1967: 1027), it is described as a compound word formed by merging the modifier úgy and the question word -e.[13] Its phonological reduction must have occurred no later than the 18th century. Example 1 illustrates its basic use as an adverb in a text from 1786:

example 1
Emlékezz meg micsoda fortélyokkal élt szeretõd? Eleinte **ugye** keveset reméltetett veled? (*Remember what trickery your lover presented in the beginning. She made you hope little then, didn't she?*) (Verseghy 1786: 247)

It can be observed in the above example that the function of *ugye* here is not only asking for confirmation, but at the same time, it calls attention to a fact, and reminds the hearer of something.

Besides its use as an adverb (dating from as early as 1585), Benkő (1967: 1027) lists two other senses of *ugye*: sentence word and filler. Its first recorded use as a sentence word dates back to 1604 and can be glossed as lám or bezzeg in Hungarian, or translated to German as fürwahr or wahrlich. As a filler, it can be translated to German as Füllwort.[14]

It is also worth mentioning that Ballagi (1867-1872) glosses "*ugye*?" and "*ugyebár*?" under the same entry as "*nemde*?" (~isn't it?), involving the concept of negation in its synonymous definition which is interesting since many languages (e.g. Germanic languages) encode questions expecting a positive reply in negated forms (e.g. German "nicht wahr?" or English "isn't it?" among other variant forms).

Based on the material in the Hungarian Historical Corpus (for instance, see examples 2 and 3), it can be hypothesized that the use of

[13] cf. the evolution of the question word/DM -e in Schirm (2011: 46-49)

[14] German translations/equivalents are provided in Benkő (1967).

ugyebár as a question word was much more common in the 18th and 19th century than it is today.

example 2
Most tehát szabadon bé-mehetek, **ugyebár**? (*I'm allowed to enter now, aren't I?*) (Szerelemhegyi 1795: 55)

example 3
Rajta magyar, készen áll már Jellacsics, Hogy hátáról címerednek bõrt hasíts! Címeredet összetépted úgyis már, S kutyabõrre szükséged van, **ugyebár**? Fékezzük meg a bitorlót álmiban. (*Up you stand, Hungarians, ready is Jellacsics, for you to cut leather from his back and embellish your armour! After all, you had destroyed you armour, and now what you are need is dog skin, is that so?*) (Sárosi 1848/1954: 184)

Besides "*ugyebár*?", Szarvas & Simonyi (1893) list "*igyé*?"and "*úgy-é*?" among the variants of *ugye*, and also provide its counterpart in Latin, which is "*siccine*?". Exemplified by example 4, *ugyé* must have also been a frequent form in the 18th century and earlier:

example 4
Az agynak táplálása, melegitése sok vért kéván. De **ugyé**, hogy az agyvelö valaha igen kitsin vala, és most egy nagy emberben igen nagyra nevekedet. … Ugyé, hogy az álom-is az ilyen spiritusnak elve ß tése miatt következik. 'S hát azt szabadé tagadni, a' mit nem látunk. (*Nurturing, warming our brain requires a lot of blood. But is that so, that the cerebrum was once tiny, and has grown big in tall man?…Is that so that dream follows from the loss of spirit, too. And, well, can we deny what we do not see?)* (Rácz 1789: 92)

Interestingly, the function of *ugye* in example 4 is not marking a question or signalling the expectation of the speaker to receive a positive reply; but, rather, it precedes different moves of an explanation; and, therefore, structures the explanation into smaller segments.

Example 5 further illustrates that *ugye* in the 18th century was already used not only to mark questions but to mark confirmation, emphasis and reassurance as well:

example 5
SOÓS MIHÁLY (utánnok kiált) : Hallá az úr, hallá! Hát már az én fiamat béveszi-é vagy nem?
PHOEBUS : Bé, bé, no!
SOÓS MIHÁLY : Hát melyik lesz elébb?
PHOEBUS : Aki utóljára olvasott verset.
SOÓS MIHÁLY (a fiának szomorúan): Ennye, megmondám, ugyé fiam! (Phoebusnak) Hallá, hallá, nagyobb az én fiam.
(SOÓS MIHÁLY (calling the others): Hear me, my Lord! Does my son get admission, or does he not?
PHOEBUS: So be it, so!
SOÓS MIHÁLY: And which one comes next?
PHOEBUS: The one that has just read out a poem.
SOÓS MIHÁLY (sadly to his son): Oh Gods, I have told you, my son, haven't I? (to Phoebus): Hear me now, hear me, greater is my son). (Nagy 1796/1964: 79)

4.3 Interim conclusion: the cline of development of the evolution of ugye into a DM

Using Traugott & Dasher's (2002) model (outlined in section 3), the following cline of development towards DM category membership can be drawn in the case of *ugye*:

úgy van-e? → *úgy-e?* → merging and phonological reduction → *ugye*: interrogative adverb → question expecting positive reply and/or re-assurance → marker of evidentiality expecting the confirmation of the validity of the content of the host utterance of the DM

5 A discourse-pragmatic analysis of *ugye* (~'is that so?') as a DM

This section overviews the discourse-pragmatic analysis of 185 *ugye* tokens used in 50 recordings of the HuComTech corpus (of which 103 are uttered by the interviewees, 82 instances by the interviewer).

5.1 Pragmatic functions of ugye

Among various approaches, we can approach the function of *ugye* from the perspective of **processing information** along the lines of

Jucker and Smith (1998), who distinguish between *reception markers* (e.g. *oh, okay*), which mark reactions to first-pair parts in adjacency pairs (e.g. statements, questions), and ***presentation markers***, which elaborate on and/or alter the information provided by the previous speaker. *Information-centred presentation markers*, such as *mondjuk* modify the unit of information itself, while ***addressee-centred presentation markers***, such as *szerintem (~'I think'), ugye* (~'is that so?', 'as you know') or *vagyis* (~'or rather') relate the information to the assumed knowledge state of the addressee.

Based on a previous contrastive study of the use of English and Hungarian DMs in mediatised political interviews (Furkó & Abuczki to appear) as well as the examples illustrating the use of *ugye* as a marker of evidentiality, it can be argued that Hungarian *ugye* is similar to English of course in terms of its role in **information management**. O*f course,* similarly to *ugye*, can be used to signal new information and/or the newsworthiness of a given utterance (cf. Furkó 2007: 105). For instance, in example 6 *ugye* is used to preface **background information** in an inserted comment; therefore, it functions as a **marker of explicitation or specification**. Concerning its prosodic features, it is preceded and followed by voice break and characterized by lower intensity (displayed in Figure 1).

example 6 (http://www.youtube.com/watch?v=t7mEXNMxjAc)
boldog élmény volt, mikor elhoztuk, **ugye** egy <egy> sharpeiről van szó, {b} és hát ilyen kis pici {l} volt, és nagyon aranyos (*It was a happy memory when we took him home, it's* ***DM_ugye*** *a shar pei, and so he was tiny like that, and very cute*) (hucomtech 016 informal)

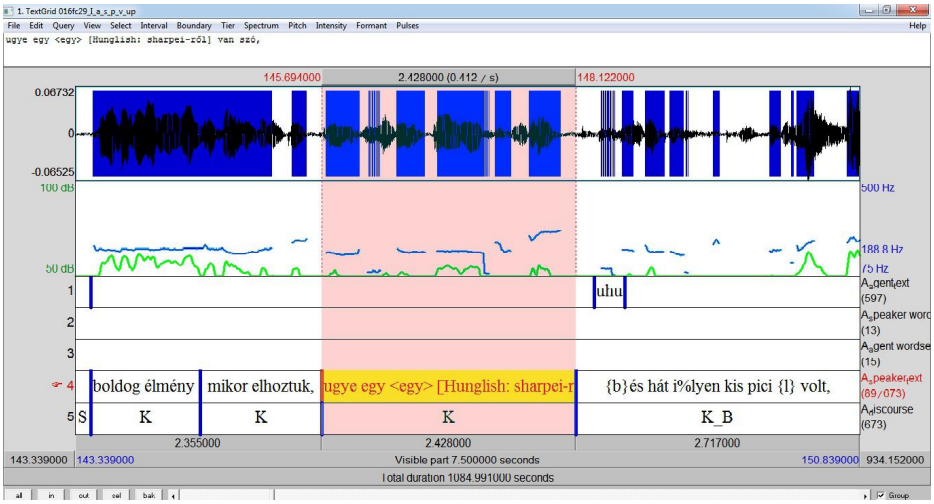

Figure 1: Acoustic analysis of example 6 in Praat

The following example (7) illustrates the **confidential use** of *ugye*, glossed as 'as you probably deduce on the basis of information or experience we have just shared'. This utterance was preceded by an unexpected segment in the simulated job interview part where the interviewer suddenly switched to speaking English and asked questions in English. This surprised the interviewee quite a lot and he could not answer in good English. So the interviewee assumes that the interviewer believes that he needs to improve his English. This assumed belief is rather specific to the context here.

> *example 7* (http://www.youtube.com/watch?v=bn4WmEi07fk)
> Addigra úgy képzelem, hogy ha nem sikerül munkát találnom, akkor párhuzamosan elkezdeném fejleszteni a nyelvtudásomat, *mer **ugye** még most angolból is hiányosságaim vannak. (*I guess, in case I won't be able to find some job by then, I'll start to work on my English then, because **DM_ugye** now I still have a lot to learn.*) (020 hucomtech informal)

In the following two examples (8 and 9) *ugye* may either **refer back in the discourse** and, **clarify previous content** or **mark the evidentiality of the information** in the inserted clause:

*example 8 (*http://youtu.be/tuOgE6mmrEs)
Különböző világok vannak, ezek tudatállapotok, és mindegyik ilyen világban, **ugye** fokozatosan haladunk felfelé, és a legfelsőbb szinteken az istenek élnek. (*There are different worlds, these are different states of mind, and in each of these worlds, **DM_ugye** we gradually move upward, and gods live at the highest level.*)

*example 9 (*http://youtu.be/ZLuKY9DAk4w)
Csak otthon hát **ugye** nyugalom kell hozzá, hogy ne zavarják az embert mondjuk egy félóráig. (*However, **DM_ugye** at home you need to have silence to do it, nobody should disturb you for at least half an hour.)*

Ugye functions as a **marker of shared knowledge, an evidentiality marker** in the utterance below (example 10) since both speakers in the interview are currently university students so they are likely to follow the changes in the system of higher education.

*example 10 (*http://youtu.be/Y1glqW3CPLQ)
Eddig Debrecenben tanultam, az Ady Endre Gimnáziumban hatosztályos képzésben, most pedig az egyetemen éppen végzős vagyok az anglisztika szakon, a BA képzésben, **ugye** ebben az új három éves képzésben. (*I've been studying in Debrecen, first in the six-year-programme of Ady Endre Grammar School, and now I'm doing my final year majoring in English BA, **DM_ugye** in this new three-year programme.*) (hucomtech 006 formal)

In the next narrative sequence[15] (example 11), the role of *ugye* is rather **fuzzy** as it can be interpreted as a combination of **evidentiality marker** or **preface of giving an excuse or explanation**.

*example 11 (*http://www.youtube.com/watch?v=OPSDK81ariU)
Aztán amikor ilyen százharminc-száznyolcvannal mentünk, akkor így elgondolkoztam, hogy lehet most kéne megállni, és nem így hajszolni lényegében így a halált, meg így kockára tenni így az életemet, de végülis **ugye** ott volt az adrenalin, és nagyon felpörgetett, és az úgy jó volt. (*Then, when we were driving at hundred-thirty, hundred-eighty kms/h, I started to think, like, that we should maybe stop now, and not, like, practically chasing death,*

[15] DMs have different functions in narrative sequences (cf. Norrick 2001).

plus, like, risking my life, but then again, there was ***DM_ugye*** *the adrenaline spinning me up a lot, and it was a great feeling.*) (hucomtech 018 informal)

As examples 12 and 13 show, *ugye* is frequently used in its **information checking** function as well in the HuComtech corpus, typically uttered in rising intonation

*example 12 (*http://youtu.be/1sd9BirHc9Q)
I, agent, 153 .s (21.)
Ő az, akiről meséltél egyszer a –
Biztos meséltem.
Most anyukádéknál van, **ugye?**
Igen, igen,
Ha jól emlékszem?
(Is he the dog you were telling me about?
I'm sure I told you.
Is he at your parents now ***DM_ugye*** *?*
Yes, yes.
If I remember well.)

*example 13 (*http://youtu.be/XGKfsZQWSuo)
Az volt a legjobb szállás, mert az egy ilyen kastélyszállószerűség volt, de nem a városba, hanem egy picit kijjebb tőle.
Hm.
De már nem jut eszembe, hogy mi a neve, de például
Ez a Győr felé, **ugye**?
Igen, igen.
(That was the best accommodation because it was like a castle, not in the city, but a little way out of it.
Hm.
But I can't remember the name, also, for instance…
Is it on the way to Győr ***DM_ugye****?*
Yes, yes.)

As shown in the situation below (example 14), the DM *ugye* also functions to **refer back to previously mentioned/given content/information**:

example 14 (http://www.youtube.com/watch?v=B_gTI332tms)

Hát a szomorú élmény az meg szintén ide kapcsolódik amikor el kellett válni tőle. Hát amikor elpusztult szegény, **ugye** a csau-csau, az előző kutya.
Ja, uhum.
Az elég rossz volt.
(Well, the sad memory is also connected to this experience. When we had to say good bye to him. When he deceased, the poor thing, ***DM_ugye*** *the chau chau, the previous dog.*
Yeah, uhum.
That was pretty bad.)

In **narrative sequences** (e.g. examples 15 and 16) *ugye* displays **story structure** and segments events and narration (cf. Norrick 2001).

*example 15 (*http://www.youtube.com/watch?v=q5KbZthmJHc)
Most a legutóbbi, amit hallottam, az is ilyen Chuck Norrisos, hogy nemtom valami filmjéből jelenet, hogy a med-- a medvével, és **ugye** hogy hogy nem is tom, hogy így alszik ott az erdőbe ilyen tábor-- mármint a medve, vagy vagy nem, Chuck-- Chuck Norris, és akkor **ugye** jön a medve, és így megtámadja, és hát akkor az is így van ilyen komoly küzdelem, aztán végülis a medve így elkezd hátrálni.
(Now the last one I've heard is a Chuck Norris one too, that, dunno, some movie of his had a scene with a b- bear, and that, the bear is like sleeping ***DM_ugye*** *in the forest or whatever at that campsite - the bear, I mean…or, or no, Chuck-- Chuck Norris, and then* ***DM_ugye*** *the bear is coming and attacking him, and, well, there is like this serious fight then, then the bear starts to back off finally.)*

*example 16 (*http://www.youtube.com/watch?v=5PFI4cw2-LI)
Ja igen, és akkor attól ijedtem meg, hogy **ugye** megyek be, és akkor apukám meg ott állt az ajtóban. És akkor attól nagyon megijedtem, mer nem számítottam rá. *(Oh yes, and then I got scared when I was* ***DM_ugye*** *entering the door, and my dad was there standing in the door. And that really scared me, 'cause I hadn't expected that to happen.)*

DM *ugye* also plays a very salient role in **explanation** (as illustrated in example 17 below).

*example 17 (*http://youtu.be/ZYXxFoghubI)
Igen, hát a krisnások például arra szoktak hivatkozni, hogy ez **ugye** benne van a izébe Bhagavad-Gítában, hogy ezt és ezt nem szabad.
Ja, ja, ja.
És ez **ugye** tényleg egy régi könyv.
*Yes, well, the Krishna people for instance usually refer to **DM_ugye** the Bhagavad Ghita or whatever, that this is mentioned in that, and you mustn't do it.*
Yeah I know.
*And this is **DM_ugye** in fact really an old book.*

5.2 Lexical co-occurrences of ugye

After mapping their roles and functions, I also aimed at mapping the lexical environment of the DM; therefore, I run a query (called Find N-gram within annotations) in ELAN on the segmented DM *ugye* tokens. The tables below (1 and 2) present the co-occurrence patterns of *ugye*, but only show those DMs and connectives that preceded *ugye* in at least two instances.

DM or connective preceding *ugye*	Number of occurrences	Proportion (%)
meg (~'and')	7	6,93
hát (~'well')	6	5,94
**mer* (~'cause')	4	3,96
akkor (~'then')	4	3,96
hogy (~'that')	4	3,96
és (~'and')	3	2,97

Table 1: A list of DMs and connectives preceding ugye in the interviewees' talk

DM or connective preceding *ugye*	Number of occurrences	Proportion (%)
hát (~'well')	6	8,96
**mer* (~'cause')	5	7,46
hogy (~'that')	4	5,97
akkor (~'then')	2	2,99
mert (~'because')	2	2,99
tehát (~'so')	2	2,99
és (~'and')	2	2,99

Table 2: A list of DMs and connectives preceding ugye in the interviewer's talk

It can be concluded from the queries that *ugye* was preceded by a DM or a connective in approximately 40% of the cases (altogether 103 tokens) in the interviewees' talk, while in the interviewer's speech *ugye* followed a DM or a connective in approximately 43% of the cases (altogether 82 tokens). Most of these *ugye*-clusters are used in explanations.

5.3 Proposed classification of the discourse-pragmatic functions of *ugye*

Taking into consideration all instances of *ugye* in the analysed parts of the HuComTech corpus, *ugye* as a DM expresses the following meanings and functions:

ugye1 used as an interrogative adverb of manner, functioning as a tag question, uttered with high pitch (the most ancient, the core meaning of *ugye*):

- ugye_1a used to signal a question asking for reassurance and/or expecting a positive reply (agreement or acceptance)
- ugye_1b used to check information

ugye2 used as a marker of evidentiality (mostly used in explanations):

- ugye_2a refers to common knowledge and can be glossed as 'as everyone knows' (impersonal use of *ugye*)
- ugye_2b is used to refer to/activate/emphasize the shared background knowledge of the speakers in the conversation (confidential use of *ugye*)
- ugye_2c indicating a rhetorical question where the speaker does not expect to receive a reply (rhetorical use of *ugye*)
- ugye_2d is uttered when the speaker mentions something that s/he assumes the other participant in the conversation already knows
- ugye_2e is sometimes used to avoid implying that someone should know the thing that you are saying but you are not certain about it

ugye3 used to express emphasis:

- ugye_3a as a marker of new information to call or evoke the hearer's attention (to pieces of new/unknown information)
- ugye_3b used to emphasize the validity of facts
- ugye_3c used to emphasize that the speaker has chosen certain words or communicated certain content/message
- ugye_3d used to introduce explanations and excuses or clarify the use of words by introducing an explanation
- ugye_3e used as a marker of story structure and new development in narratives

6 Attempts at disambiguating discourse functions

As discussed in sections 1 and 2, the scrutinised lexical item provides information about discourse structure and dialog acts which helps listeners process the utterance and helps speakers to organize conversational moves, such as transition relevant places (TRPs). From all their possible functions, two salient functions of the selected DM will be described by multiple multimodal features in order to increase its naturalness in spoken language generation and to enhance its semi-automatic meaning disambiguation. This goal is crucial in natural language processing since the proper generation and understanding of such frequent lexical items based on multimodal description might

be a component of interactive voice response systems, especially for producing and interpreting communicative messages.

6.1 Material and feature extraction methods

I have selected two functions of *ugye* to analyse in detail because (1) they are salient[16] functions of the given DM and (2) the realisations of these two different functions seem to be relatively easy to distinguish by relying on only observable, measurable and machine-extractable features. I only tagged prototypical members of the categories and excluded borderline cases with ambiguous and highly mixed functions. The specification of this strict tagging might result in better precision results in classification attempts.

Low-level prosodic features such as min, max and mean pitch and intensity, as well as temporal features (durations of DMs and preceding silences) were extracted from the segmented sound files (.wav) using Praat (Boersma & Weenink 2007) and Prosogram scripts (Szekrényes, Csipkés & Oravecz 2011, Szekrényes to appear). The values of the F0 range and intensity range of each DM were calculated in Microsoft Excel. Speaker normalization can also be achieved using Z-scores. Lexical features concerning lexical-co-occurrences and position (in the utterance) were extracted from the textgrid files of the annotations. Nonverbal-visual features of the speaker's behaviour can be extracted from the manually-performed video annotations of the recordings and can be automatically queried with the ELAN software (using the command Find overlapping label of 'wordseg' in the 'gaze' or/and 'facial expression' tiers.).

On the other hand, visual features can also be machine-extracted from the video files with relatively high reliability using facial expression recognition software, such as the FaceReader software of Noldus.[17] Several single layer search options and multiple layer search options will be performed to identify the common features of the prototypical uses of the three items.

---

[16] Criteria for saliency was the high frequency of the function assigned during a series of questionnaires (described in detail in Abuczki to appear) as well as my qualitative observations about its frequency in the HuComTech corpus.

[17] You can read about FaceReader and request a free trial version of it at http://www.noldus.com/human-behavior-research/products/facereader.

Finally, the machine-extractable features will be organized in order of prominence into a decision tree which quite reliably (with at least 80% precision, depending on the end node) distinguishes the two salient uses of the given DM.

The material of this empirical case study is comprised of twenty-two informal recordings18 of the HuComTech corpus. This cross-checked subcorpus contains 70 tokens of *ugye*. The two salient major functions of *ugye* (~'is that so?') to be distinguished were (1) marker of explanation and evidentiality and (2) marker of question and asking for approval and reassurance. The category of marker of explanation and evidentiality (abbreviated as EXPLN) includes tokens used during acts of explanation, including evidentiality markers and markers of shared knowledge, since all these two categories overlapped in the responses of my questionnaire informants. Example 18 below is a protoypical use of *ugye* as a marker of explanation and evidentiality. Markers of question and asking for approval and reassurance (abbreviated as QSTN) include information-checking and facilitating (tag) questions that ask for reassurance. Example 19 presents a canonical use of *ugye* as a marker of question (aking for reassurance and checking information).

marker of explanation and evidentiality:

example 18
"BA-s hallgató vagyok **ugye** ezt a 3 éves képzést csinálom". ('*I'm a BA student DM_ugye I do this three-year course.*') (HuComTech, 006_I);

marker of question:

example 19
Pestre fogsz költözni, **ugye**? ('*You're moving to Pest, DM_ugye?*') (HuComTech, 085_I)

In my sample analysed 23 tokens fall into the category of EXPL and 15 into the category of QSTN out of the 70 tokens of *ugye*.

[18] I selected these twenty-two recordings because on the one hand, I wanted to have equal distribution of male and female speakers in this subcorpus; on the other hand, all the audio, video, pragmatic and functional DM annotations of these recordings had been checked at/by the time of writing this chapter.

6.2 The relation of discourse function and manual gesticulation

As an initial assumption, I had expected to find correspondences between discourse functions and hand movements. During the queries of manual gestures, I considered any kind of hand movement (and any handshape type other than the default handshape type of the current speaker) as manual gesticulation. I queried the relation of hand gesticulation and each of the salient functions of DMs one by one in separate queries (with the 'Find overlapping labels' command), and then joined them in contingency tables (see Table 3) for statistical analysis in SPSS 19. 0.

The difference in the frequency and extent of hand gestures is considerable in the case of *ugye* where explanations are frequently, while questions are very rarely performed with simultaneous manual gesturing (shown in Table 3 and Figure 2).

Function_of_ugye * Manual_gesticulation Crosstabulation				
		Manual_gesticulation		Total
		present	absent	
Function_of_ugye	Explanation	16	7	23
	Question	1	14	15
Total		17	21	38

Table 3: Ugye (~'is that so?'): the relation of function and hand movement

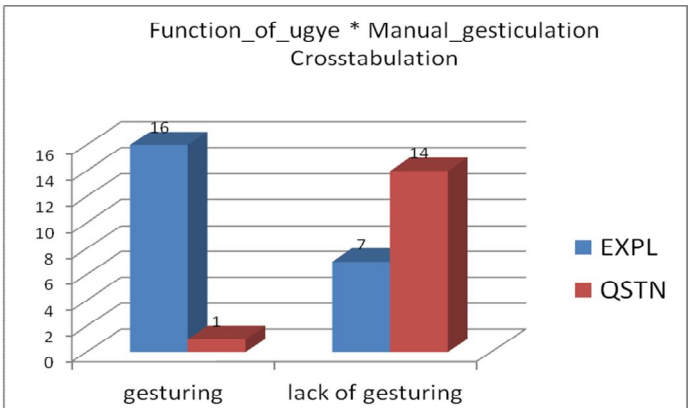

Figure 2: Ugye (~'is that so?'): the relation of function and hand movement

The results suggest that explanation is typically accompanied by the movements of the hands, as presented in Figure 3.

Figure 3: Explanation and simultaneous hand movements

As a qualitative observation, I have found that new lexical-semantic information (new in comparison to what has just been uttered previously in the context of the conversation) are most often brought into the domain of discourse during the turn-keep and topic elaboration discourse segments. Pieces of novel information are almost always marked, usually not verbally, but by nonverbal means, very often accompanied by heavy gesturing, especially bilateral open hand manual gesturing.

6.3 The relation of discourse function and gaze direction

Similarly to the role of facial expression, I also expected to find a relation between discourse function and gaze direction; therefore, I queried the distribution of gaze direction during ugye_QSTN and ugye_EXPL. The following figure (Figure 4) presents the typical forward gaze direction type (that I wanted to confirm by the queries) displayed right after a tag question:

Figure 4: Ugye_question and the simultaneous cessation of manual gesturing, gaze at the other participant/next speaker selection

As the bar chart in Figure 5 shows, while the distribution of gaze direction types during explanations is rather varied and balanced, the overarchingly most frequent gaze direction type during tag questions is forwards, in other words, eye contact (given the seating position of the speakers facing each other) which marks giving the floor over to the listener.

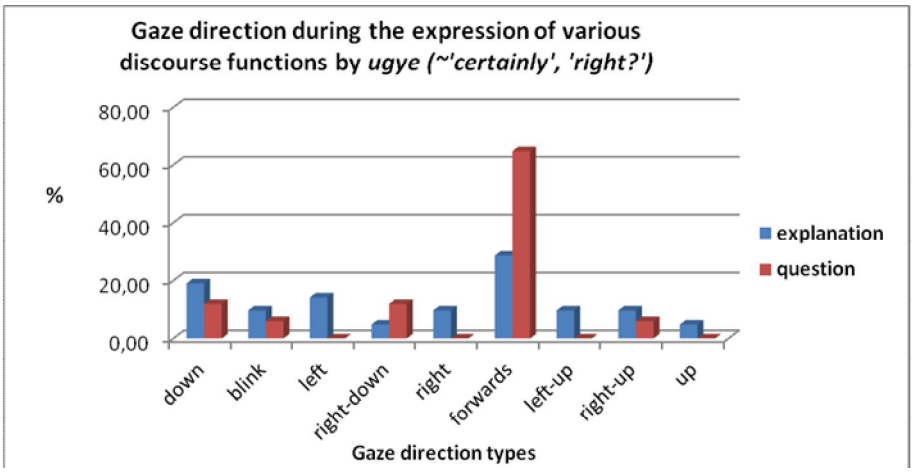

Figure 5: The distribution of gaze direction types during explanations and tag questions

In sum, the distinctive nonverbal-visual properties of the functions of *ugye* (~'is that so?') based on queries and qualitative analyses are the following:

hand movement:

- the presence of manual gesticulation is typical in explanations
- while the absence of hand movement characterizes questions

gaze direction:

- shifting gaze of the speaker is typical of explanation
- forward gaze/eye contact is common in questions

6.4 The relation of discourse function and preceding silence

The present section describes the query results of automatic segmentation of the sound files into silent and sounding parts in Praat. This analysis followed the segmentation of DMs and it was performed with the aim to test if DMs are predominantly preceded by silence.

It was found that *ugye* tokens rarely occur in utterance-initial position and are rarely preceded by silence. Instead, they are typically placed in the middle or end of clauses and utterances, and therefore, they are preceded by sounding parts. As it can be read in Table 4 and Figure 6, the difference between the categories of explanation and question are again not significant (p>0,05) in terms of the frequency of preceding pauses.

Function_of_ugye * Silence_preceding_the_DM Crosstabulation				
		Preceding_silence		Total
		present	absent	
Function_of_ugye	Explanation	4	19	23
	Question	3	12	15
Total		7	31	38

Table 4: Ugye (~'is that so?'): the relation of function and preceding silence

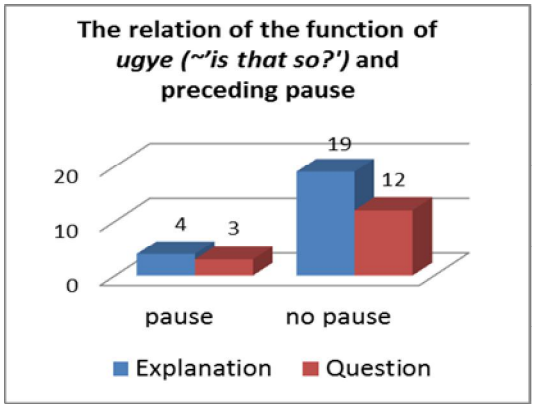

Figure 6: Ugye (~'is that so?'): the relation of function and preceding silence

6.6 The relation of discourse function and F0

After the segmentation and functional annotation of DMs, a Praat script (Boersma & Weenink 2007) was run in order to measure min, max and mean values of F0, intensity and duration of the individual DM tokens performing the two most salient functions, and save them in a spreadsheet file. In a few cases, F0 values could not be defined, therefore, DM tokens with undefined F0 values were dropped out of the analysis. Only *ugye* tokens with explanation (EXPL) and question (QST) functions were analysed in terms of these prosodic parameters. Owing to the physical and speech production differences of the two sexes, F0 values of female (Figure 7) and male speakers (Figure 8) were queried and analysed separately.

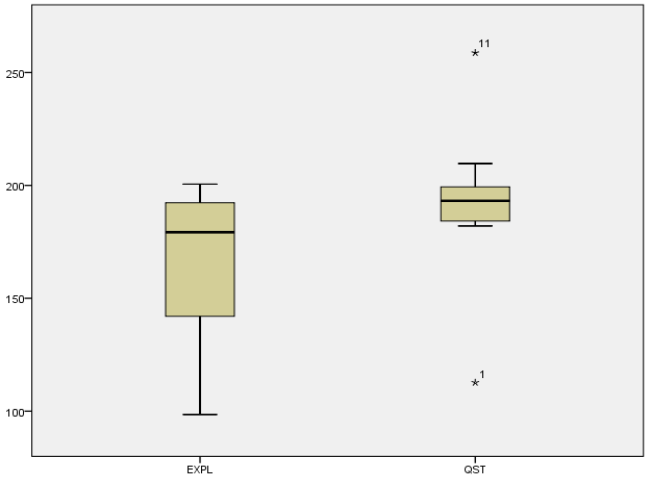

Figure 7: Distribution of mean F0 of ugye with different functions among female speakers

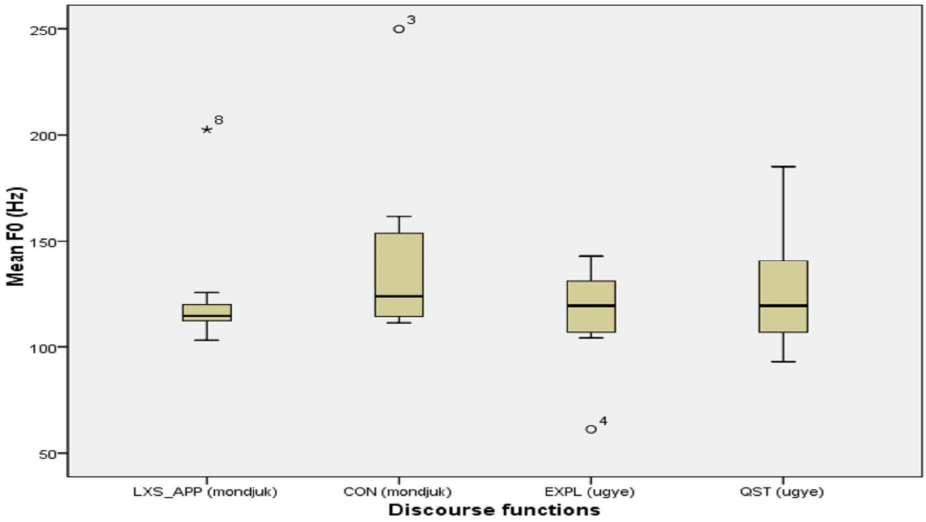

Figure 8: Distribution of mean F0 of ugye with different functions among male speakers

The difference between the mean F0 values of DMs expressing different functions is not significant either among male or female speakers. However, interestingly, the standard distribution of F0 values in different functions differs among the five male and five female speakers. Among male speakers the range of F0 values is larger in DMs with

question functions than among female speakers. The frequency data of female speakers shows that they realize *ugye* (~'as you know') tokens used in explanations with lower fundamental frequency. It is not surprising that the majority of the speakers (especially female speakers) pronounce *ugye* in question function with relatively higher pitch than in explanatory sequences in assertions (presented in Figure 7) since Gussenhoven & Chen (2000) claim that high pitch, as a suprasegmental/paralinguistic feature, universally marks dependence and questioning. It is surprising though that the F0 data of male speakers (presented in Figure 8) do not reflect this tendency.

6.7 Conclusions: Modelling multimodal features in a decision tree

The most typical and reliable defining feature distinguishing different functions of the analysed DM is the simultaneous activation or the lack of manual gesticulation. Based on several single layer queries (most of them presented, some are not), the following sets of features were found to describe the expressions of the salient functions of the scrutinised DM:

ugye (~'as you know') marking explanation: presence of manual gesticulation, lower mean F0 than in questions, various positions (but mostly utterance-internal);

ugye (~'isn't it?') in questions: higher mean F0 than the mean F0 of its host unit as well as that of explanation, rising intonation, larger F0 range than during ugye_EXPL, absence of hand movement, mostly gaze forward (eye contact with the conversation partner, marking giving the floor over to him/her with the question), various positions (rarely utterance-initial, but mostly utterance-final).

After identifying the individual prominent features that distinguish the two types of the selected item, I tried to organize the machine-extractable features into a decision tree[19]. Naturally, besides machine-detectable cues, the speech act (illocutionary force) of the

[19] My considerations for modelling multimodal features in decision tress as can be found in Abuczki (2012a: 335-342). I have earlier developed a decision tree that similarly relies on multimodal features (such as gaze direction, manual gesturing, duration of segments and silences) and distinguishes silent moments in conversation based on contextual (previous and simultaneous multimodal cues) as either markers of turn ends (transition relevant places, cf. Sacks 1992) or turn-internal pauses (Bódog, Abuczki & Németh T. 2011: 245-246).

DM's host unit as well as its lexical environment also have impact on its functional interpretation but these are not considered here due to the difficulty of their automatic analysis. I experimented with various parameters and threshold values of each machine-extractable feature that may best distinguish the expression of two functions. ELAN enables multi-level, hierarchical, embedded queries in any combination of the tiers and labels, called Multiple Layer Search. I used this search option in a great number of variations. My ultimate goal was to identify parameters and threshold values that best distinguish categories and divide tokens into two categories in such a way that at least 80% of the tokens of the divided categories fall into the same category; therefore, the reliability/precision of the decision tree would be at least 80%, depending on which node the decision ends in (the later/the further down it ends, the more reliable the query is, between 80% and 100%).

Figure 9 illustrates the distinctive features of two uses of *ugye*, as a marker of explanation (EXPL) as opposed to a marker of question (QSTN), organized into a decision tree that can distinguish the two salient functions of a word with at least 80% precision[20] (depending on the end note of the decision process).

[20] Manually checked results.

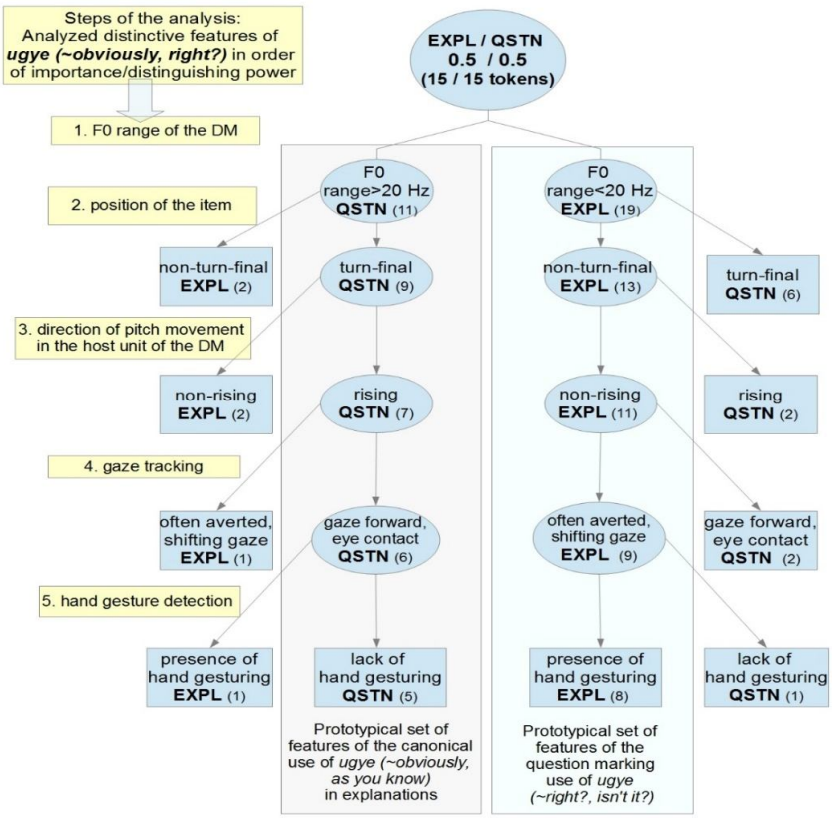

Figure 9: A possible decision tree distinguishing between salient functions of ugye with multimodal features

The prototypical multimodal features of *ugye* (true for 80% of the cases21) as marker of question are the following (in order of importance):

- F0 range of the DM22 > 20 Hz
- turn-final position
- rising or upward intonation[23] (in the DM or its host unit)

21 80% of the analysed cases in the HuComTech corpus.

22 Automatically extracted by using a Praat script (Boersma & Weenink 2007)

23 The annotation of pitch movement was automatically performed (into five categories) and then manually checked.

forward gaze direction of the speaker, making eye contact with the listener (longer than during explanations), signalling the intention to give the floor over to the listener
lack of hang gesturing, indicating the end of speaking and speech planning

In contrast, the prototypical multimodal features[24] of *ugye* as a marker of explanation are as follows:

- F0 range of the DM < 20 Hz
- turn-initial or turn-internal position
- stagnant, fall or descending intonation (in the DM or its host unit)
- shifting gaze direction of the speaker (forward, sideways, up)
- hand gestures performed

It must be noted as a limitation of the generalizability of the results that the decision trees was modelled based on the data of only 30 *ugye* tokens (15 tokens of EXPL, 15 tokens of QSTN); however, it may serve as a springboard for further theoretical modelling.

7 Conclusions

It can be concluded that specific combinations of multimodal characteristics together help the interlocutors in the disambiguation of both the DM and discourse in general. I managed to identify two sets of multimodal surface cues that distinguish two salient functions of each of *ugye* with 80% precision. To distinguish between the two salient uses of *ugye*, I also involved the analysis of intonation and included F0 range and pitch movement direction. Furthermore, sequential features such as the position of the DM also seemed to play a role in disambiguation (following F0 range), but the distribution of facial expression labels and DM duration values were not significantly different between the two categories.

[24] Each set of features describes at least 80% of the members of the analysed category.

References

Abuczki, Á. (2012): A conversation-analytical study on turn-giving cues: end-of-turn prediction. In: Esposito, A., Esposito, A.M., Vinciarelli, A., Hoffmann, R. & Müller, V.C.: *Behavioural Cognitive Systems, LNCS 7403*. Berlin: Springer-Verlag, 335-342.

Abuczki, Á. (monograph/to appear): *A Core/Periphery Approach to the Functional Spectrum of Discourse Markers in Multimodal Contex*t (doctoral dissertation).

Aijmer, K. (2002): *English Discourse Particles: Evidence from a Corpus*. Amsterdam & Philadelphia: John Benjamins.

Ballagi M. (szerk.) (1867-1872): *A magyar nyelv teljes szótára.* Első rész. Budapest: Franklin Társulat (MNyTSz.)

Bell, D. (1998): Cancellative discourse markers: a core/periphery approach. *Pragmatics* 8 (4), 515-541.

Benkő, L. (szerk.) (1967): *A magyar nyelv történeti-etimológiai szótára.* 1. kötet. A-Gy. Budapest: Akadémiai Kiadó.

Benkő, L. (szerk.) (1992): *A magyar nyelv történeti nyelvtana II/1. A kései ómagyar kor. Morfematika.* Budapest: Akadémiai Kiadó.

Bódog A., Abuczki Á. & Németh T. E. (2011): A multimodális pragmatikai annotáció jelentősége a számítógépes nyelvészetben. In: *VIII. Magyar Számítógépes Nyelvészeti Konferencia.* Szeged: Szegedi Tudományegyetem, 240-251.

Boersma P. & Weenink, D. (2007): *Praat: doing phonetics by computer 5.0.02.* University of Amsterdam: Institute of Phonetic Sciences. http://www.praat.org (last accessed: 05.05.2014)

Brugman, H. & Russel, A. (2004): Annotating multi-media / multimodal resources with elan. In: Lino, M., Xavier, M., Ferreire, F., Costa, R. & Silva, R. (eds.): *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC).* Lisbon: Portugal, 2065-2068.

Erman, B. & Kotsinas, U. (1993): Pragmaticalization: the case of ba' and you know. *Studier i Modern Språkvetenskap. Acta Universitatis Stockholmiensis, New Series* 10, 76-93.

Fraser, B. (1999): What are discourse markers? *Journal of Pragmatics* 31, 931-952.

Fraser, B. (2006): Towards a theory of discourse markers. In: Fischer, K. (ed.): *Approaches to Discourse Particles*. Oxford & Amsterdam: Elsevier, 189-204.

Fraser, B. (2009): Topic orientation markers. *Journal of Pragmatics* 41, 892-898.

Furkó, P.B. (2007): *The Pragmatic Marker-Discourse Marker Dichotomy Reconsidered – the Case of Well and Of Course*. Debrecen: Kossuth Egyetemi Kiadó.

Furkó, B.P. & Abuczki, Á. (2014): English discourse markers in mediatised political interviews. *Brno Studies in English* 40 (1).

Gussenhoven, C. & Chen, A.-J. (2000): Universal and language-specific effects in the perception of question intonation. *ICSLP 6. Vol. II*. Beijing, 91-94.

Gyuris, B. (2008): A diskurzus-partikulák formális vizsgálata felé. In: Kiefer F. (szerk.): *Strukturális magyar nyelvtan IV. A lexikon szerkezete*. Budapest: Akadémiai Kiadó, 639-682.

Hansen, M.-B. M. (2006): A dynamic polysemy approach to the lexical semantics of discourse markers. In: Fischer, K. (ed.): *Approaches to Discourse Particles*. Oxford & Amsterdam: Elsevier, 21-42.

Hopper, P. & Traugott, E.C. (1994, 2003): *Grammaticalization*. Cambridge: Cambridge University Press.

Ittzés, N. (ed.) (2006): *A magyar nyelv nagyszótára*. II. Budapest: MTA Nyelvtudományi Intézete.

Jucker, A.H. & Smith S.W. (1998): And people just you know like 'wow': Discourse markers as negotiating strategies. In: Jucker, A. H. & Ziv, Y. (eds.): *Discourse Markers: Descriptions and Theory*. Amsterdam & Philadelphia: John Benjamins, 171-203.

Kugler N. (2000): A módosítószó. In: Keszler B. (szerk.): *Magyar Grammatika*. Budapest: Nemzeti Tankönyvkiadó, 298-302.

Lewis, D. (2006): Discourse markers in English: a discourse-pragmatic view. In: Fischer, K. (ed.): *Approaches to Discourse Particles*. Oxford & Amsterdam: Elsevier, 43-59.

Norrick, N.R. (2001): Discourse markers in oral narrative. *Journal of Pragmatics* 33, 849-878.

Péteri, A. (2002): *Abtönungspartikeln im deutsch-ungarischen Sprachvergleich*. Budapest: ELTE.

Pusztai F., Juhász J., Gerstner K. & J. Soltész K. (eds.) (2003): *Magyar értelmező kéziszótár*. Budapest: Akadémiai Kiadó.

Sacks, H. (1992): *Lectures on Conversation*. Oxford: Blackwell.

Schiffrin, D. (1987): *Discourse Markers*. Cambridge: Cambridge University Press.

Schirm A. (2009): Diskurzusjelölők a parlamenti beszédekben. In: Kukorelli Katalin (szerk.): *Hatékony nyelvi, idegen nyelvi és szakmai kommunikáció interkulturális környezetben.* Dunaújváros: Dunaújvárosi Főiskola, 168-175.

Schirm A. (2011): *Diskurzusjelölők funkciói: a hát, az -e és a vajon elemek története és jelenkori szinkrón státusa alapján*. Doktori értekezés. Szeged: SZTE.

Szarvas G. & Simonyi Zs. (szerk.) (1893): *Magyar Nyelvtörténeti Szótár*. Budapest: Hornyánszky Viktor Akadémiai Könyvkereskedése (MNyTSz.)

Szekrényes, I. (to appear): Annotation and interpretation of prosodic data in the HuComTech corpus for multimodal user interfaces. *Journal on Multimodal User Interfaces*. Berlin & Heidelberg: Springer.

Szekrényes I., Csipkés L. & Oravecz Cs. (2011): A HuComTech-korpusz és -adatbázis számítógépes feldolgozási lehetőségei. Automatikus prozódiai annotáció. In: Tanács A. & Vincze V. (szerk.): *VIII. Magyar Számítógépes Nyelvészeti Konferencia.* Szeged: JATEPress, 190-198.

Thanh, N. (2006.): From procedural meaning to processing requirement. In: Fischer, K. (ed.): *Approaches to Discourse Particles.* Oxford & Amsterdam: Elsevier, 167-188.

Traugott, E.G. (1995): The role of the development of discourse markers in a theory of grammaticalization. *Paper given at the 12th International Conference on Historical Linguistics.* Manchester, 13-18, August, 1995.

Traugott, E.G. & Dasher, R.B. (2002): *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

Sources of diachronic data

Hungarian Historical Corpus (Magyar Történeti Szövegtár): http://www.nytud.hu/hhc/

Nagy I. (1964): *Debreceni diákirodalom a felvilágosodás korában. Lakadalmi játék*. Budapest: Akadémiai Kiadó, 79.

Rácz S. (1789): *A physiologiának rövid sommája.* Pest: Patzkó, 92.

Sárosi Gy. (1954): *Kisebb költeményei, prózai munkái és levelezése*. Budapest: Akadémiai Kiadó, 184.

Szerelemhegyi A. (1795): *A jó-tévö szarándok, vagy-is A tsörgö sapka.* Pest: Trattner, 55.

Verseghy F. (1982): *Verseghy Ferenc Kiadatlan írásai. 1. Erkölcsi levelek*. Szolnok: Verseghy Ferenc Megyei Könyvtár, 247.

Source of synchronic data

HuComTech corpus: http://lingua.arts.unideb.hu/hucomtech-database/ (last accessed: 08.22.2013)

Dr. Ágnes Abuczki
MTA-DE Research Group for Theoretical Linguistics
Pf. 47
H–4010 Debrecen
abuczki.agnes@gmail.com