Why *almost* and *almost* are not even approximately the same: The diachronic semantics of approximatives in Hungarian[1]

Tamás Halm
halm.tamas@gmail.com
www.tamashalm.com
Research Institute for Linguistics (Hungarian Academy of Sciences)
Pázmány Péter Catholic University

In our paper, we will explore the diachronic semantics of two almost-approximators in Hungarian: *majdnem* ('almost-M') and *szinte* ('almost-S'). We will show that there is a neat division of labour between these two: *majdnem* encodes epistemic vagueness (possible worlds, cf. Sadock 1981, Nouwen 2006 a.o) whereas *szinte* encodes scalar vagueness (precision standards, cf. Penka 2005, Sauerland & Stateva 2007, Amaral & del Prete 2010, Greenberg & Ronen 2013 a.o.), a situation similar to that observed in Russian (čut' ne and počti, cf. Kagan and Wolf 2015). We will show that this can be straightforwardly derived from the grammaticalization trajectories of the two approximators: *majdnem* derives from the adverb *majd* 'soon' (and the expletive negator *nem* 'no') and *szinte* derives from an adverb originally meaning 'by outward appearance, superficially'. These two processes both exhibit semantic bleaching and the semanticization of originally pragmatically inferred information (Eckardt 2006). With the help of corpus data, we will also track and explain in terms of competing grammars (Niyogi 2002) how, in certain environments, *majdnem* is in the process of crowding *szinte* out.

1. Introduction

In this paper, I will explore the diachronic formal semantics of two *almost*-approximators in Hungarian: *majdnem* ('almost-M') and *szinte* ('almost-S'). From a synchronic perspective, I will show that there is a neat division of labour between these two: *majdnem* encodes epistemic vagueness (possible worlds, cf. Sadock 1981, Nouwen 2006 a.o) whereas szinte encodes scalar vagueness (scales of varying granularity or a contextually given precision standard, cf. Penka 2005, Sauerland & Stateva 2007, Amaral & del Prete 2010, Greenberg & Ronen 2013 a.o.), a situation similar to that observed in Russian (*čut' ne* and *počti*, cf. Kagan and Wolf 2015). These findings lend further support to the dualistic view of vagueness.

From a diachronic perspective, I will show that this can be straightforwardly derived from the gram-maticalization trajectories of the two approximators: *majdnem* derives from the adverb *majd* 'soon' (and the expletive negator *nem* 'no') and *szinte* derives from an adverb originally meaning 'by outward appear-ance, superficially' (Simonyi 1881, Simonyi 1888). These two processes both exhibit semantic bleaching and the semanticization of originally pragmatically inferred information (Eckardt 2006). With the help of corpus data, I will also track and explain in terms of competing grammars (Niyogi 2002) how, in certain environments, *majdnem* is in the process of crowding *szinte* out.

2. Data

Hungarian has two words which are usually rendered as *almost* in English: *majdnem* és *szinte*. These have typically been regarded as stylistic alternatives (with *szinte* considered the more refined alternative), however, a closer look at various environments reveals a more complex picture. In some environments, the

1 I wish to express my gratitude to Ágnes Bende-Farkas, Katalin É. Kiss and the participants of SinFonIJA 12 (Brno), Nyelvelmélet és diakrónia 4 (Budapest) and FoDS4 (Jerusalem) conferences. My research was supported by a Postdoctoral Grant of the Hungarian Academy of Sciences (PPD 031/2017), Project Grant 129921 of the National Research Foundation of Hungary and Grant TKP2020-NKA-11 of the Ministry for Innovation and Technology of Hungary.

two approximators can be used interchangeably. The figures below denote attestations in the Hungarian National Corpus (Váradi, Oravecz & Sass 2014):

(1) a. *János és Máté majdnem egyidősek.* (HNC=40)
John and Matthew almost-M same.age$_{Adj}$:PL
b. *János és Máté szinte egyidősek.* (HNC=48)
John and Matthew almost-S same.age$_{Adj}$:PL
'John and Matthew are almost the same age.'

Both (1a) and (1b) are completely felicitous and they have identical truth conditions. The figures in parentheses denote the number of attestations of the relevant structures (*majdnem egyidősek* and *szinte egyidősek* 'almost the same age'). In other environments, however, a striking pattern emerges:

(2) a. *Majdnem első lettem a futóversenyen.* (HNC=21)
almost-M first became:1SG the running.competition:on
b. *#Szinte első lettem a futóversenyen.* (HNC=0)[2]
almost-S first became:1SG the running.competition:on
'I almost came first in the running competition.'

While (2a) is perfectly natural, (2b) is strongly marked: although it is grammatical in terms of syntax, it is semantically anomalous. If uttered, (2b) would probably met with a rejoinder along the lines: *'Well, you cannot be* almost the first*, you are either first or not first, there is no in-between.'*

A similar contrast can be observed below:

(3) a. *Majdnem pap lettem.* (HNC=5)
almost-M priest become.PAST.1SG
b. *#Szinte pap lettem.* (HNC=0)[3]
almost-S priest become.PAST.1SG
'I almost became a priest.'

(3b) is anomalous as priesthood is typically conceived of as a binary condition: one is either a priest or not, there are no grades in between.[4]

The inverse pattern is observable in (4) below:

(4) a. *#A mozaikok majdnem véletlenül kerültek Ravennába.* (HNC=6)
the mosaics almost-M by.accident got Ravenna:into
b. *A mozaikok szinte véletlenül kerültek Ravennába.* (HNC=136)
the mosaics almost-S by.accident got Ravenna:into
'The mosaics ended up in Ravenna almost by accident.'

(4a) can only receive a felicitous reading in the rather absurd scenario where there was a devious plan to get the mosaics to Ravenna in an accidental manner, but in the end, this plan failed, and the mosaics got to Ravenna in a non-accidental manner. Such a scenario is difficult to reconcile with our world knowledge and this makes (4a) infelicitous (or at least strongly marked). A similar contrast can be observed below:

2 Includes: *első* 'first', *bajnok* 'champion', *aranyérmes*, 'gold medallist', *dobogós* 'one who earned the right to stand on the podium reserved for the first three', *listavezető* 'first in the list of contenders', *pápa* 'pope', *elnök* 'president', *király* 'king', *miniszterelnök* 'prime minister'

3 Includes: *szerzetes* 'monk', *apáca* 'nun', *katona* 'soldier'

4 Theologically and liturgically speaking, there do in fact exist stages of priesthood: lector, acolyte, deacon, priest and bishop. However, in everyday parlance, only the latter two are regarded as priests.

(5) a. *#Majdnem testvérek vagyunk.* (HNC=2)
almost.M siblings be-1PL
b. *Szinte testvérek vagyunk.* (HNC=27)
almost-S siblings be-1PL
'We are almost brothers.'

(5b) is felicitous on the reading 'We are as close to each other as siblings typically are.' With (5a), this reading is not accessible and the sentence is infelicitous. (More precisely, it would only be felicitous under an unlikely and convoluted scenario where we somehow almost ended up as brothers but did not in the end.) A similar pattern is shown below:

(6) a. *#10 év alatt majdnem soha nem láttam a főnököt.* (HNC=10)
ten year under almost-S never not see.PAST.1SG the boss.ACC
b. *Tíz év alatt szinte soha nem láttam a főnököt.* (HNC=1.061)
ten year under almost-S never not see.PAST.1SG the boss.ACC
'In my ten years at the company, I almost never saw the boss.'

(6a) is only felicitous under the following convoluted scenario: This is my last day at work. So far, I have never met the boss. And then, unexpectedly, the boss steps into my office, so on my very last working day, I do meet the boss. I almost managed never to see the boss, but I did in the end.

Finally, the two approximators pattern differently with regard to universal quantifiers and free-choice items:

(7) a. *#Ezt a feladatot majdnem bárki meg tudja oldani.* (HNC = 10)
this.ACC the task.ACC almost-M anyone PRT can.3SG solve.INF
b. *Ezt a feladatot szinte bárki meg tudja oldani.* (HNC=181)[5]
this.ACC the task.ACC almost-S anyone PRT can.3SG solve.INF
'Almost anyone can solve this task.'
c. *Ezt a feladatot majdnem mindenki meg tudja oldani.* (HNC=806)
this.ACC the task.ACC almost-M everyone PRT can.3SG solve.INF
d. *Ezt a feladatot szinte mindenki meg tudja oldani.* (HNC=3922)
this.ACC the task.ACC almost-S everyone PRT can.3SG solve.INF
'Almost everyone can solve this task.'

While a universal quantifier can be felicitously combined with either *majdnem* 'almost-M' or *szinte* 'almost-S', a free-choice item is infelicitious with *majdnem* 'almost-M'.

3. Previous proposals

Some instances of the non-interchangeability of *majdnem* and *szinte* have been noted by some authors. Halm (2016a) mentions the contrast between *#majdnem bárki* 'almost anyone' and *szinte bárki* 'almost anyone'. Dékány & Csirmaz (2018) point out that *majdnem* is easier to combine with numerals than *szinte* (see discussion later). Dékány & Csirmaz (2018) also claim that *majdnem elég* 'almost-M enough' is more felicitous than *szinte elég* 'almost-S enough'; this, however, appears to be not borne out by the facts: the Hungarian National Corpus gives a similar number of attestations for both versions (24 vs. 24). In addition to these empirical observations, Piñón (2008) discussess the semantics of *majdnem* 'almost-M' in passim and sketches a possible-world analysis not dissimilar to ours, without, however, discussing *szinte* 'almost-S' or the contrast between *majdnem* and *szinte*.

5 As far as the other free-choice paradigm is concerned (cf. Halm 2016b), the results are the following:
(i) *#majdnem akárki* (HNC = 0)
szinte akárki (HNC = 2)

In my proposal, I will follow the general view in the literature (see references above) that in the case of approximators such as majdnem or szinte, two meaning components are to be distinguished:[6]

(8) *János majdnem/szinte két méter magas.*
John almost two meter high
'John is almost two meters high.'
PROXIMAL component: John is close to being two meters high.
POLAR component: John is not two meters high.

While the semantics of the polar component is relatively straightforward, there is a wider variety of proposals when it comes to modelling the proximal component. However, two broad directions emerge: epistemic vagueness vs. scalar vagueness.

- Epistemic vagueness (possible worlds): While in w_0 (our current world), John is not two meters high, there is alternative world w_1 close to w_0 in which John is two meters high. (Sadock 1981, Nouwen 2006 a.o.)
- Scalar vagueness: under the current, contextually given precision standard it is untrue that John is two meters high; there is, however, a slightly laxer precision standard under which it is true that John is two meters high.[7] (Sauerland & Stateva 2007, Greenberg & Ronen 2013 a.o.)

Slightly modifying the formulas proposed by Greenberg & Ronen (2013)[8], the two approaches can be represented as follows:

- EPISTEMIC VAGUENESS (one proposition, alternative possible worlds, distance between possible worlds):

(9) a. POLAR $\neg p_{w0}$
b. PROXIMAL $\exists w_1 \in S_{ALT}(w).\ close_s(w_1,w_0) \wedge p_{w1}$

- SCALAR VAGUENESS (alternative propositions, one possible world, distance between precision standards):

(10) a. POLAR $\neg p_{preC,w0}$
b. PROXIMAL $\exists pre' \in S_{ALT}(preC).\ close_s(pre',preC) \wedge p_{pre',w0}$

4. Synchronic analysis

In this section, I propose a model of the behaviour of the two approximators from a synchronic perspective. I argue that within the grammar of Modern Hungarian, *majdnem* 'almost-M' is an approximator defined in epistemic vagueness terms: (Sadock 1981, Nouwen 2006 a.o.):

6 I assume that directionality (the less-than/before effect with numerals and temporal or spatial expressions) is a by-product of polarity (cf. Penka 2005 and Amaral and Del Prete 2010 a.o.)

7 Other authors conceptualize scalar vagueness without referring to standards of precision, making the simpler assumption that the proposition p modified by the approximator has a set of alternatives, and these alternatives can be ordered into a scale wrt to one another and p (Penka 2005, Nouwen 2006, Amaral & Del Prete 2013). Greenberg & Ronen (2013) use scales of precision with regard to *more or less* (English) and *paxot o yoter* (Modern Hebrew), but only assume the existence of scalar alternatives when it comes to modelling *almost* (English) and *kim' at* (Modern Hebrew).

8 (9) is the specification of the general formula proposed by Greenberg & Ronen (2013:6) on the basis of Greenberg & Ronen (2013:6-7). (10) is a formalization of the model proposed by Sauerland & Stateva (2007) expressed in the general formula proposed by Greenberg & Ronen (2013:6).

(11) *majdnem* p

a. POLAR $\neg p_{w0}$

b. PROXIMAL $\exists w_1 \in S_{ALT}(w).\ close_s(w_1,w_0) \wedge p_{w1}$

Whereas *szinte* 'almost-S' is an approximator defined in scalar vagueness terms (Sauerland and Stateva 2007, Greenber and Ronen 2013 a.o.):

(12) *szinte* p

a. POLAR $\neg p_{preC,w0}$

b. PROXIMAL $\exists pre' \in S_{ALT}(preC).\ close_s(pre',preC) \wedge p_{pre',w0}$

This model provides a very good fit with the empirical patterns observed in Section 2. The contrast observed in (2) falls out readily: (2a) containing *majdnem* 'almost-M' is felicitous under a readily accessible scenario where the speaker ended up second but would have ended up first in a slightly different possible world, e.g. one in which he did not happen to lose his shoes in the last metres before the finish line. (2b) with *szinte* 'almost-S' is infelicitious though: there does not exist a meaningful precision standard, however lax, under which someone who lost the race may be truthfully asserted as having won the race.

Sentences (1a) and (1b) are both felicitous. If John is only 3 days older than Matthew, than it is the case that under a somewhat laxer precision standard, they are of the same age, a situation covered by sentence (1b) with *szinte* 'majdnem-S'. Also, if two possible worlds only differ in whether the age differential between John and Matthew is 0 days or 3 days, then these two possible worlds are likely to be close under any reasonable closeness metric, and thus, (1a) with *majdnem* is also fully felicitous.

Sentence (4b) with *szinte* 'almost-S' descibes felicitously a situation where the manner in which the mosaics got into Ravenna was not fully accidental strictly speaking, but can be characterized as accidental under a slightly vaguer precision standard. (4a) with *majdnem* 'almost-S' is degraded since the counterfactual reading is not compatible with our world knowledge: it is difficult to conceive of a plan where the explicit purpose was to reach a well-defined, concrete end result in an accidental manner: one cannot achieve something accidental by design.

The proposal to model *majdnem* 'almost-M' in epistemic vagueness terms and *szinte* 'almost-S' in scalar vagueness terms is also supported by the observation that counterfactual readings are only available with *majdnem* 'almost-M'. Consider:

(13) a. *János majdnem pontosan érkezett,…*
John almost-M on.time arrive.PAST.3SG
'John almost arrived on time,…

i. *aligegy percet késett.*
just one minute.ACC be.late.PAST.3SG
he was only a minute late.'

ii. *de pár kilométerrel a cél előtt lerobbant a kocsija,*
but couple km.INS the goal before stopped.working the car.his
but a couple of kilometres before his destination, his car broke down,
és végül egy jó órát késett.
and eventually a good hour.ACC be.late.PAST.3SG
and he ended up being more than an hour late.'

b. *János szinte pontosan érkezett,…*
John almost-S on.time arrive.PAST.3SG
'John almost arrived on time,…

i. *alig egy percet késett.*
just one minute.ACC be.late.PAST.3SG

he was only a minute late.'

ii. *#de pár kilométerrel a cél előtt lerobbant a kocsija,*
but couple km.INS the goal before stopped.working the car.his
but a couple of kilometres before his destination, his car broke down,
és végül egy jó órát késett.
and eventually a good hour.ACC be.late.PAST.3SG
and he ended up being more than an hour late.'

The continuation (ii) is only compatible with a counterfactual reading and indeed, as expected, this continuation is felicitous with *majdnem* 'almost-M' but infelicitous with *szinte* 'almost-S'.

This neat divison of labour between *majdnem* 'almost in epistemic vagueness terms' and *szinte* 'almost in scalar vagueness terms' lend considerable support to a dualistic view of vagueness in the semantics of approximators. It should be noted that Kagan and Wolf (2015) have reached a similar conclusion with regard to Russian, where *čut' ne* 'almost' is an approximator in epistemic vagueness terms and *počti* 'almost' is an approximator in scalar vagueness terms.

5. Diachronic analysis – *majdnem*

In this section, we will have a close look at the diachronic formal semantics (and syntax) of *majdnem* 'almost-M'. I will argue that the synchronic semantics of *majdnem* can be neatly derived from its diachronic sources and trajectory.

At first look, in Modern Hungarian, *majdnem* 'almost-M' seems to be some sort of compound made up of the temporal adverb *majd* 'soon' and the negator *nem* 'not'. However, it appears unlikely that synchronically, the meaning of *majdnem* 'almost-M' could be derived compositionally from these elements: 'it is almost the case that p' and 'it will soon not be the case that p' seem to be distant propositions.

It is important to note, though, that the adverbial *majd* also has a second reading, restricted to fossilized emphatic statements and (for some speakers) to numerals:

(14) a. *Majd(nem) elfelejtettem.*
almost-M PRT.forget.PAST.1SG
'I almost forgot.'

b. *Majd(nem) elájultam.*
almost-M PRT.faint.PAST.1SG
'I almost fainted (with admiration/of fear).'

c. *Majd(nem) felrobbantam.*
almost-M PRT.explode.PAST.1SG
'I almost exploded (with rage).'

(15) *János majd(nem) két méter magas.*
John almost-M two meter high
'John is almost two meters tall.'

Here, *majd* on its own (without *nem* 'not') is interpreted as an almost-approximator: 'almost-M'. Furthermore, archaically and in very formal registers, a third variant is also attested in Modern Hungarian: *majdhogynem* 'almost-M', which looks like an amalgamation of *majd* 'soon', *hogy* 'that' and *nem* 'not':

(16) *A gyengébb tanulók számára majdhogynem*
the weaker students for almost-M
megoldhatatlanok voltak az érettségi feladatok.
unsolvable.PL were the final.exam tasks
'For the weaker students, the tasks in the final exam were almost impossible to solve.'

That is, in addition to the productive *majdnem* 'almost-M', there are also two marginal variants: *majd* 'almost-M' and *majdhogynem* 'almost-M'. As is often the case with fossils, it is reasonable to assume that these marginal variants are remnants from the earlier stages of the diachronic trajectory that gave us Modern Hungarian *majdnem* 'almost-M'.

Indeed, as discussed in the historical linguistics literature (cf. the Historical Dictionary of Hungarian 2:819 and references therein), *majd* 'almost-M' is attested a good 250 years earlier than *majdnem* 'almost-M':

(17) *zertelen* *tezȯn* *magaual,* *mayd* *el* *vezti* *feiet*
unrestrained do.3SG with.himself almost PRT lose.3SG his.head.ACC
'He fails to restrain himself, he almost loses his head.' Guary Codex (before 1508, 027)

It is easy to see how the approximator *majd* 'almost-M' came about from the temporal adverb *majd* 'soon'. Consider:

(18) a. *Majd elájulok.*
soon PRT.faint.1SG
'I will faint soon.'[9]
b. *Majd elájulok.*
soon PRT.faint.1SG
'I am almost fainting.'

(18a) asserts that there is a w_1 such that it is a temporally close continuaton world of w_0 where the speaker is fainting. Since w_1 and w_0 are temporally close, they are also close qua possible worlds under any closeness metric. There is also a scalar implicature to the effect that the speaker is not fainting as of now.

(19) *majd* 'soon'

TEMPORAL	w_1 is a continuation world of w_0	(asserted)
PROXIMAL	$\exists w_1 \in S_{ALT}(w).\ close_s(w_1,w_0) \wedge p_{w1}$	(entailed[10])
POLAR	$\neg p_{w0}$	(scalar implicature)

The reinterpretation of *majd* 'soon' into *majd* 'almost-M' involved semantic bleaching (a well-know hallmark of grammaticalization): the temporal meaning component was lost. At the same time, the proximal component was reinforced (from logically entailed to asserted), and the polar component became part of the semantic meaning (as opposed to being a scalar implicature):

(20) *majd* 'almost-M'

PROXIMAL	$\exists w_1 \in S_{ALT}(w).\ close_s(w_1,w_0) \wedge p_{w1}$	(asserted)
POLAR	$\neg p_{w0}$	(entailed[11])

Note that such semanticization of originally pragmatically inferred information has also been described as a typical feature of grammaticalization (Eckardt 2006).

---

9 Note that in Hungarian, finite verb forms are either marked with a past tense morpheme or they carry no tense morphology. In the latter case, the verb is underspecified in terms of tense and is ambiguous between a present tense or a future tense reading. Hence the ambiguity of *Elájulok.* 'PRT.faint.1SG' between 'I am fainting.' and 'I will faint.'

10 Logical entailment of the temporal component.

11 The precise status of the polar component in approximators in subject to considerable debate (cf. Roberts 2011 and Horn 2002, 2011 and references therein). Here I adopt Horn's (2002) proposal that the polar component is semantically entailed but assertorically inert.

Importantly for our purposes, this grammaticalization process also explains why *majd* 'almost-M' (and its descendant *majdnem* 'almost-M') are approximators defined in epistemic terms with a possible worlds semantics.

What remains to be determined is the appearance and amalgamation of *nem* 'not', in other words, how *majdnem* 'almost-M' emerged from *majd* 'almost-M'. As Simonyi (1888) and others (cf. Historical Dictionary of Hungarian 2:819) have claimed, *majdnem* is the amalgamation of *majd* 'almost' and the expletive negator *nem* 'not', however, they stepped short of actually mapping out a grammaticalization pathway. The earliest attestations are the following:

(21) *a' mit én tegnap néktek mondottam, majd hogy nem elég arra,*
what I yesterday to.you say.past.1sg, almost-M that not enough for.that
hogy boldogul lenne dolgotok
that happily would.be your.affairs
'What I told you yesterday is almost enough to make you content with the state of your affairs.' (Ferenc Földi. 1790. Erkölcs-könyvecske, 44)

(22) *majdnem = vix non, fere, ferme, propermodum, paene* (Ferenc Verseghy. 1816. Analytica institutionum linguae hungaricae.)

After taking a closer look at the relevant data in the Hungarian Historical Corpus[12], two important patterns emerge. The first is that the appearance of *majdhogynem* (spelling variant: *majd hogy nem*) 'almost-M' preceded the appearance of *majdnem* 'almost-M'. The second is that for an extended period, *majd hogy nem* 'almost-M, literally: almost-that-not' was in competition with *majd* hogy 'almost-M, literally almost-that':

(23) a. *Az én szívem pedig majd hogy meg nem hasadt.* (1852)
the my heart.1SG then almost that PRT not split
'My heart almost broke.'
b. *Sára asszony egészen megváltozott, a szive majd hogy meghasadt.* (1892)
Sarah aunt completely PRT.changed the heart.3SG almost that PRT.split
'Aunt Sarah changed completely, her hear almost broke.'

Based on these observations, we can hypothesise the following grammaticalization pathway for 'almost-M': *majd* -> *majdhogy* (lit. almost-that) -> *majdhogynem* (lit. almost-that-not) -> *majdnem* (lit. almost-not). The locus of the first grammaticalization step was a structure where majd 'almost-M' was adjacent to the complementizer hogy 'that', exemplified below:

(24) *(Majd hátra esik) Ni! ni! az Ördög vigye el,* (1793)
almost back falls well well the devil take.IMP PRT
majd hogy a' kórság belém nem áll!
almost that the disease 1SG.into not stands
'(He almost falls to his back.) There, there! May the Devil take it, I am almost struck down by the disease!'

As a first step of analyzing the syntax of sentences such as (24), note that in these sentences, *majd* 'almost-M' patterns with a family of speaker-oriented modal/evidential discourse particles such as *éppen* 'exactly, just in time':

(25) *Éppen, hogy elértem a vonatot.*

12 Late 18th to late 20th century, 30 million word tokens.

exactly that PRT.reach.PST.1SG the train.ACC
'I reached the train just in the nick of time.'

Several models have been proposed for sentences such as (25) above. Kenesei (1992) argues convincingly against a two-clause (main clause – subordinate clause) analysis and proposes that the adverbial particle occupies Spec,CP:

(26) [CP *éppen* [C' *hogy* [TP *elértem* *a* *vonatot.*]]]
exactly that PRT.reach.PST.1SG the train.ACC
'I reached the train just in the nick of time.'

The monoclausal analysis has achieved consensus, however, there are different proposals as to the position of the adverbial particle. Kenesei (2002) has argued that it occupies Spec,AdvSP (where AdvS stands for sentence adverbial):

(27) [AdvSP *éppen* [AdvS' *hogy* [TP *elértem* *a* *vonatot.*]]]
exactly that PRT.reach.PST.1SG the train.ACC
'I reached the train just in the nick of time.'

É. Kiss (2010) proposes a model where the particle occupies the head position of a speech act phrase (SAP):

(28) [SAP SPEAKER [SAP' *éppen* [CP *hogy* [TP *elértem* *a* *vonatot.*]]]
exactly that PRT.reach.PST.1SG the train.ACC
'I reached the train just in the nick of time.'

Our proposed analysis is compatible with any of the above proposals. For concreteness and simplicity, we adopt Kenesei (1992) and analyze (24) as follows:

(29) [CP' *majd* [C' *hogy* [TP *a* *kórság* *belém* *nem* *áll.*]]]
almost that the disease 1SG.into not stands
'I am almost struck down by the disease.'

As support for this analysis, note that the only conceivable alternative, a biclausal analysis, is easy to exclude as the copula can never intervene between *majd* 'almost-M' and the complementizer *hogy* 'that' (here as well, *majd* 'almost-M' patterns with éppen 'exactly'):

(30) a. **Majd volt, hogy el nem estem.*
almost was that PRT not fall.PST.1SG
intended: 'I almost fell.'
b. **Éppen volt, hogy elértem a vonatot.*
exactly was that PRT.catch.PST.1SG the train.ACC
intended: 'I caught the train just in the nick of time.'

The grammaticalization process involved the reinterpretation of the two adjacent elements, *majd* 'almost-M' and *hogy* 'that' as a single adverbial:

(31) a. [CP' *majd* [C' *hogy* [TP *meghasad* *a* *szívem.*]]]
almost that PRT.split.3SG the heart.1SG
'My heart is almost breaking.'
b. [TP [AdvP *majdhogy*] [TP *meghasad* *a* *szívem*]]]

almost PRT.split.3SG the heart.1SG
'My heart is almost breaking.'

The *éppen, hogy* 'exactly' -> *éppenhogy* 'exactly' reinterpretation probably proceeded along similar lines. The clearest sign that the reinterpretation has indeed taken place is that in Modern Hungarian, both *majdhogy* 'almost-M' and *éppenhogy* 'exactly' can freely appear clause-internally:

(32) a. *A derekam majdhogy letört.* (1943)
the back.1SG almost PRT.broke
'My back almost broke.'
b. *A kvóták éppenhogy Magyarország érdekében születtek.* (2018)
the quotas exactly Hungary interest.3SG.in were.born
'The quota system was explicitly designed to serve the interests of Hungary.'

As usual, this amalgamation into a single element was followed only with some delay by orthography and for a period of time, the amalgamated *majdhogy* continued to be spelled as *majd hogy*.

It should also be noted that in Hungarian, as well as in many other languages, complementizer-drop is widely attested under certain circumstances. This means that an underlying *majd hogy* construction may emerge as a clause-inital *majd* construction:

(33) a. [CP *majd* [C' ~~*hogy*~~ [TP *meghasad a szívem.*]]]
almost that PRT.split.3SG the heart.1SG
'My heart is almost breaking.'
b. [CP *éppen* [C' ~~*hogy*~~ [TP *jókor érkeztél.*]]]
exactly that at.good.time arrive.PST.2SG
'You arrived just at the right time.'
c. [CP *csak* [C' ~~*hogy*~~ [TP *megérkeztél.*]]]
only that PRT.arrive.PST.2SG
'You did arrive at long last.'

Turning to the semantically apparently superfluous *nem* 'not' element, we should note first that in Hungarian, expletive negation is widely attested in mirative contexts such as in wh-exclamatives or so-called surprise negation sentences (cf. Halm and Huszár 2021):

(34) a. *(hogy) mik meg nem történnek manapság!*
that what.PL PRT not happen.3PL these.days
'What (surprising) things happen these days!'
b. *(hát) nem elfelejtettem a PIN-kódomat!?*
well not PRT.forget.PST.1SG the PIN-code.1SG
'I forgot my PIN code (unexpectedly)!'

It is reasonable to assume that such an expletive *nem* 'not' could appear together with approximatives in sentences with a mirative flavour, as is indeed attested by the example discussed in (24), reproduced here for convenience as (35):

(35) *(Majd hátra esik) Ni! ni! az Ördög vigye el,*
almost back falls well well the devil take.IMP PRT
majd hogy a' kórság belém nem áll! (1793)
almost that the disease 1SG.into not stands
'(He almost falls to his back.) There, there! May the Devil take it, I am almost struck down by the disease!'

The reinterpretation of *majdhogy* 'almost-M' and *nem* 'not' as a single compound *majdhogynem* 'almost-M' was facilitated by the fact that the expletive negator is semantically vacuous (i.e., it does not encode negation on the truth-conditional, semantic level). The locus of reinterpretation must have been environments where these two elements were juxtaposed (in the absence of a verbal particle), exemplified below:

(36) *Minden sorában egy szív megreped;*
every line.3SG.in a heart PRT.breaks
Könyétől könyve majd hogy nem csepeg. (1845)
tear.3SG.from book.3SG almost-M that not drips
'In every line of his, a heart breaks,
His book is almost dripping with his tears.'

Whether *majdhogynem* came about in two consecutive steps (*majd hogy* -> *majdhogy* and *majdhogy nem* -> *majdhogynem*) or in a single step (*majd hogy nem* -> *majdhogynem*) probably cannot be answered with certainty, especially as it is perfectly possible that the two processes took place in parallel fashion:

3 STEPS:

(37) a. [CP *majd* [C' *hogy* [TP *nem csepeg*]]]
almost that not drips
b. [TP [AdvP *majdhogy*] [TP *nem csepeg*]]
almost not drips
c. [TP [AdvP *majdhogynem*] [TP *csepeg*]]
almost drips
'It is almost dripping.'

2 STEPS:

(38) a. [CP *majd* [C' *hogy* [TP *nem csepeg*]]]
almost that not drips
b. [TP [AdvP *majdhogynem*] [TP *csepeg*]]
almost drips
'It is almost dripping.'

As far as the emergence of *majdnem* 'almost-M' is concerned, there are two hypotheses to consider. One may entertain the possibility that it was derived from *majdhogynem* 'almost-M' via the drop of the element *hogy*. However, it seems unlikely that an element would be simply dropped from the middle of a word, even if the morphological makeup of the word is still accessible to some extent. A more plausible scenario is that it derives from *majd ~~hogy~~ nem* 'almost that not', that is, from structures where the complementizer was phonologically null due to that-drop (cf. (29) above):

(39) a. [CP *majd* [C' *~~hogy~~* [TP *nem csepeg*]]]
almost that not drips
b. [TP [AdvP *majdnem*] [TP *csepeg*]]
almost drips
'It is almost dripping.'

After the reinterpretation, *majdnem* 'almost-M' and *majdhogynem* 'almost-M' are monomorphemic, and the syllable *nem* has no independent meaning or function. This is illustrated by the fact that an expletive negator can freely appear together with them:

(40) *Majdhogynem el nem zavartak.*
almost PRT not send.away.PAST.3PL
'They almost sent me away.'

(41) *Forraljuk addig, amíg a víz majdnem el nem fogy.*
boil.IMP that.to until the water almost PRT not diminishes
'Boil the water until it is almost all gone.'

It should be noted that the grammaticalization pathway described above seems to be closely mirrored by other similar elements such as *csak(hogy)nem*[13] 'almost, archaic', *alighogynem*[14] 'almost, archaic', and *alig(ha)nem*[15] 'most probably, archaic'. Consider:

csak(hogy)nem 'almost, archaic':

(42) a. [CP *csak* [C' *hogy* [TP *nem csepeg*]]]
only that not drips
'It is barely not dripping.' (note that here the negation is not expletive)
b. [TP [AdvP *csakhogy*] [TP *nem csepeg*]]
except.that not drips
'Except that it is not dripping.'
c. [TP [AdvP *csakhogynem*] [TP *csepeg*]]
almost drips
'It is almost dripping.'

(43) a. [CP *csak* [C' ~~*hogy*~~ [TP *nem csepeg*]]]
only that not drips
'It is barely not dripping.'
b. [TP [AdvP *csaknem*] [TP *csepeg*]]
almost drips
'It is almost dripping.'

alighogynem 'almost, archaic':

---

[13] Providing an exhaustive answer to the question whether *csaknem* is an epistemic or scalar approximator is beyond our scope here. Nevertheless, using the test of FCI-modification shows that *csaknem* patterns with *majdnem* 'almost-M':
(i) *csaknem bárki* (HNC=1)
almost anyone
intended: 'almost anyone'
(ii) *csaknem mindenki* (HNC=176)
almost everyone
'almost everyone'
This, together with the diachronic pathway sketched above, suggests that *csaknem* is an epistemic approximator.

[14] Consider:
(i) *Hanem még a pártján lévőket is alighogynem czinkostársaknak mondja.* (1907)
rather even the party.3SG.on be.PCP.PL.ACC too almost accomplices.DAT says
'Rather, he goes as far as to almost characterize his supporters as his accomplices.'

[15] Consider:
(i) *Közvetlenül a front mögött szolgált, alignem munkásosztaga volt.* (1917)
directly the front behind served most.probably labour.platoon.3SG was
'He served directly behind the front, most probably, he was heading a labour platoon.'

(44) a. [CP *alig* [C' *hogy* [TP *nem* *csepeg*]]]
barely that not drips
'It is barely not dripping.'

b. [TP [AdvP *alighogynem*] [TP *csepeg*]]
almost drips
'It is almost dripping.'

alig(ha)nem 'most probably, archaic':

(45) a. [CP *alig* [C' *ha* [TP *nem* *csepeg*]]]
barely if not drips
'If it is not dripping, then it is only by a small margin that it is not dripping.' = 'It is most probably dripping.' (note that here the negation is not expletive)
b. [TP [AdvP *alighanem*] [TP *csepeg*]]
most.probably drips
'It is most probably dripping.'

(46) a. [CP *alig* [C' ~~*ha*~~ [TP *nem* *csepeg*]]]
barely if not drips
'If it is not dripping, then it is only by a small margin that it is not dripping.' = 'It is most probably dripping.' (note that here the negation is not expletive)
b. [TP [AdvP *alignem*] [TP *csepeg*]]
most.probably drips
'It is most probably dripping.'

Clarifying the details of these pathways has to be left for further research.

6. Diachronic analysis - *szinte*

As has been pointed out by Simonyi (1881) among others (cf. the Historical Dictionary of Hungarian 3:760 and references therein), *szinte* 'almost-S' derives from the Old and Middle Hungarian word *szín* 'outer appearance, surface', combined with a locative suffix *-t-* and the suffix *-e* (analyzed as either a lative suffix or a possessedness suffix). In Modern Hungarian, this sense of *szín* has almost completely disappeared[16], surviving only in a handful of fossils such as:

(47) *a* *víz* *színe*
the water surface.3SG
'the surface of the water'
(48) *szín-lel*
outer.appearance-verbalizer
'to pretend, to create and keep up a false appearance'

In Old Hungarian, *szinte* meant 'by appearance, by superficial similarity'. The first attestation of *szinte* 'almost-S' is from Early Middle Hungarian:

(49) *olian* *kemenioṇ […]* *mint* *ẓinte* *a* *ẓaraz foḷd* *az* *o̧* *labai* *alat*
so hard as almost the dry land the he foot.3SG.PL under

[16] In Modern Hungarian, *szín* means colour, a meaning clearly related to but distinct from the one discussed in the main text.

'Almost as hard as the dry land under his feet.' (Debrecen Codex 1519, 177)

It is easy to recognize how the grammaticalization process probably unfolded, bringing about a reinterpretation from 'by appearance, by superficial similarity' to 'by a laxer standard of precision'. Consider:

(50) a. *szinte* *teljesen* *egészséges*
by.appearance completely healthy
'completely healthy by the look of it'
b. *szinte* *teljesen* *egészséges*
almost completely healthy
'almost completely healthy'

If someone appears completely healthy upon cursory visual inspection, it follows that she is completely healthy by a laxer standard of precision. Additionally, there is a scalar implicature that she is not healthy by a stricter standard of precision: if she were, we would expect the speaker to say so, instead of making a less informative statement. More formally, *szinte* 'by physical appearance, by superficial similarity' can be represented as:

(51) *szinte* 'by physical appearance, by superficial similarity'

SIMILARITY	p_{w0} by physical appearance	(asserted)
PROXIMITY	$\exists pre' \in S_{ALT}(preC).\ close_s(pre', preC) \wedge p_{pre',w0}$	(entailed)
POLAR	$\neg p_{w0}$	(scalar implicature)

Szinte 'almost-S' can be represented as:

(52) *szinte* 'almost-S'

PROXIMITY	$\exists pre' \in S_{ALT}(preC).\ close_s(pre', preC) \wedge p_{pre',w0}$	(asserted)
POLAR	$\neg p_{w0}$	(entailed[17])

This reinterpretation involved semantic bleaching (a well-known hallmark of grammaticalization): the temporal meaning component was lost. At the same time, the proximal component was reinforced (from logically entailed to asserted), and the polar component became part of the semantic meaning (as opposed to being a scalar implicature). Note that such semanticization of originally pragmatically inferred information has also been described as a typical feature of grammaticalization (Eckardt 2006).

Archaically and dialectally, the variant *szint(e)hogy* 'almost-S' is also attested. The diachrony of this form can be analyzed in similar fashion to *majdhogy* 'almost-M' (see above). Consider:

(53) [CP *Szinte,* [C' *hogy* [TP *dagadt* *a* *mellök*
almost that swelled the breast.3PL
az *egymás* *biztatásától.*]]][18]
the each.other encouragement.3SG.from
'They were almost swelling with confidence having encouraged each other.'

(54) *Megvetett* *az* *övéhez* *képest* [AdvP *szinthogy*] *értéktelen* *gépjárművem* *[…]* *miatt.*[19]
despised the his.ALL compared almost valueless car.1SG because.of
'He despised me because of my car, which was almost valueless compared to his.'

[17] Cf. footnote 11.
[18] Váth János: Ég a Papnádas. 1929. In: Váth János: Balatoni levegőben. Balatonfüred: Balatoni Szövetség.
[19] https://kiszamolo.hu/gazdagnak-latszani-mindenaron/comment-page-2/, blogpost, dated: 2017 October 22nd, accessed: 2020 June 21st.

Interestingly, expletive negation is not attested with szinte 'almost-S'. This might be connected to the absence of counterfactual readings. As a consequence, the forms **szint(e)nem* and **szint(e)hogynem* are also unattested.

7. An ongoing development

In this section, I discuss an ongoing development. In almost+numeral environments, both *majdnem* and *szinte* are attested:

(55) *Majdnem két méter magas vagyok, és hozzá száz kilót nyomok*
almost-M two meter high be.1SG and to.that hundred kilo.ACC weigh.1SG
'I am almost two metres high and on top of that, I weigh a hundred kilos.'

(56) *Danny igazi ember-hegy: szinte két méter magas és hatalmas izmai vannak.*
Danny real man-mountain almost-S two meter high and huge muscle.3SG.PL be.3PL
'Danny is a real mountain of a man: he is almost two metres high and has huge muscles.'

However, as has been pointed out by Dékány & Csirmaz (2018), for most speakers, *szinte* is marked here, something which our proposal does not explain. Adding a diachronic dimension, evidence from the Hungarian Historical Corpus shows that in those environments where *majdnem* and *szinte* compete (such as almost+numeral constructions), *majdnem* is in the gradual process of crowding *szinte* out, with the change showing the well-known logistic curve (or S-curve, cf. Kroch 1990; Niyogi & Berwick 1997):

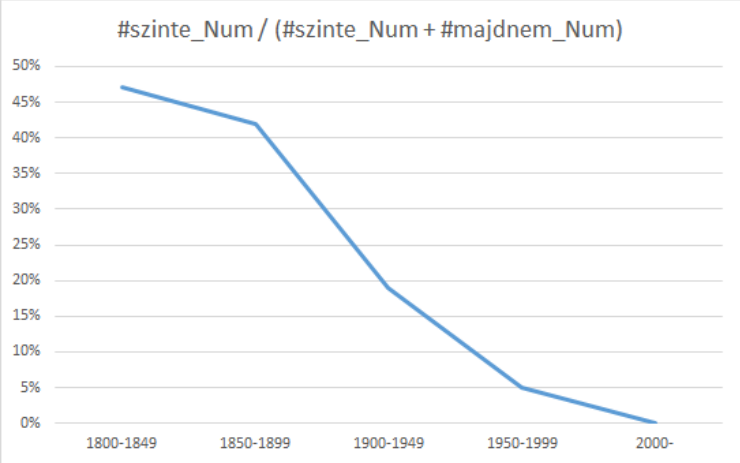

Chart 1: The proportion of *szinte* in almost+numeral constructions

An explanation of this observation can be offered in terms of Niyogi's (2002) model of language acquisition and competing grammars. According to this framework, the direction of change between two competing grammars depends on the relative frequency of the following environments:

- environments where both *szinte* 'almost-S' and *majdnem* 'almost-M' can be used
- environments where only *szinte* 'almost-S' can be used
- environments where only *majdnem* 'almost-M' can be used

While in principle, considering their semantics, both *majdnem* 'almost-M' and *szinte* 'almost-S' can be freely combined with numerals, szinte (implying a relaxation of precision standards) is less felicitous and more marked with numeral+unit combinations that suggest precision:

(57) a. *János majdnem két méter magas.*

John almost-M two meter high
'John is almost two metres high.'

b. *János szinte két méter magas.*
John almost-S two meter high
'John is almost two metres high.'

c. *János majdnem 190 centi magas.*
John almost-M 190 cm high
'John is almost 190 cms high.'

d. *??János szinte 190 centi magas.*
John almost-S 190 cm high
'John is almost 190 cms high.'

e. *János majdnem 191 centi magas.*
John almost-M 191 cm high
'John is almost 191 cms high.'

f. *#János szinte 191 centi magas.*
John almost-S 191 cm high
'John is almost 191 cms high.'

This means that *majdnem* has a competitive edge in terms of language acquisition in these environments. Through successive cycles of language acquisition across generations, this advantage causes majdnem to crowd out szinte in these competitive environments.

8. Conclusion

In this paper, I explored the formal semantics of two almost-approximators in Hungarian, *majdnem* and *szinte*, from a synchronic and diachronic perspective. I argued that the traditional view which regards these too elements as stylistic alternatives is incorrect: there exist environments where only one of them can be used felicitously. I pointed out that this distribution pattern falls out naturally if we assume that *majdnem* is an almost-approximator expressing epistemic vagueness, whereas *szinte* is an almost-approximator expressing scalar vagueness. From a historical perspective, I showed that the synchronic formal semantics of these approximators derives neatly from the trajectories of their respective grammaticalization pathways. *Majdnem* derives from the temporal adverb *majd* 'later, soon' and developed into an approximator defined in possible world terms. Szinte derives from an adverb meaning 'by appearance, by superificial similarity', and developed into an approximator defined in scalar vagueness terms.

REFERENCES

Amaral, Patricia and Fabio Del Prete. 2010. Approximating the limit: the interaction between quasi 'almost' and some temporal connectives in Italian. *Linguistics and Philosophy* 33(2), 51-115.

Benkő, Loránd (ed.) 1967-1976. *A magyar nyelv történeti-etimológiai szótára. [Historical Dictionary of Hungarian.]*

Greenberg, Yael and Moria Ronen. 2013. *Three approximators which are almost / more or less / be-gadol the same.* In Nora Boneh (ed.): Proceedings of IATL28.

Dékány, Éva and Anikó Csirmaz. 2018. Numerals and quantifiers. In Gábor Alberti and Tibor Laczkó (eds.): *Syntax of Hungarian: Nouns and noun phrases.* Amsterdam University Press, 1044-1149.

Eckardt, Regine. 2006. Meaning change in grammaticalization: an enquiry into semantic reanalysis. Oxford University Press.

É. Kiss, Katalin. 2010. Valószínűleg, hogy román kontaktushatás. In É. Kiss, Katalin és Hegedűs Attila (szerk.): Nyelvelmélet és kontaktológia. 223-238.

Halm, Tamás. 2016a. The syntactic position and quantificational force of FCIs in Hungarian. *Acta Linguistica Academica* 63(2), 241-276.

Halm 2016b. The Grammar of Free-Choice Items in Hungarian. PhD dissertation, Pázmány Péter Catholic University.

Halm, Tamás and Anna Huszár. 2021. Expletive Negation in Exclamatives - Evidence from Hungarian. *Acta Linguistica Hungarica.*

Horn, Laurence R. 2002. Assertoric inertia and NPI licensing. *Proceedings of the Annual Meeting of the Chicago Linguistic Society. Volume 38, Part Two: The Panels.* University of Chicago.

Horn, Laurence R. 2011. Almost forever. In Etsuyo Yuasa, Tista Bagchi and Katharine Beals (eds.): *Pragmatics and autolexical grammar: In honor of Jerry Sadock.* John Benjamins, 3-21.

Piñón, Christopher. 2008. Weak and strong accomplishments. In Katalin É. Kiss (ed.): *Event structure and the left periphery: Studies on Hungarian.* Springer, 91–106.

Kagan, Olga and Lavi Wolf. 2015. Gradability versus Counterfactuality: Almost in English and Russian. *Proceedings of IATL 30.*

Kenesei, István 1992. A HKM mondathatározóval. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan I. Mondattan.* Akadémiai Kiadó. Budapest. 680-1.

Kenesei István 2002. *Hányféle igazság van?* Magyar Nyelv 98: 39-49.

Kroch, Anthony. 1990. Reflexes of grammar in patterns of language change. Language Variation and Change 1. 199–244.

Morzycki, Marcin. 2001. Almost and its kin, across categories. *Proceedings of SALT11,* 306-325.

Niyogi, Partha 2002. The Computational Study of Diachronic Linguistics. In: David Lightfoot (ed.): *Syntactic Effects of Morphological Change.* Cambridge University Press, Cambridge.

Niyogi, Partha & Robert C. Berwick. 1997. Evolutionary consequences of language learning. Linguistics and Philosophy 20. 697–719.

Nouwen, Rick. 2006. Remarks on the polar orientation of almost. *Linguistics in the Netherlands* 23(1), 162-173.

Penka, Doris. 2005. Almost: a test? In Paul Dekker and Michael Franke (eds.): *Proceedings of the 15th Amsterdam Colloquium,* 179-184.

Rapp, Irene and Arnim von Stechow. 1999. Fast 'almost' and the visibility parameter for functional adverbs. *Journal of Semantics 16,* 149-204.

Roberts, Craige. 2011. Only: A case study in projective meaning. *Baltic International Yearbook of Cognition, Logic and Communication 6(1)*, 1-59.

Sadock, Jerrold. 1981. Almost. In Peter Cole (ed.): *Radical Pragmatics.* New York: Academic Press, 257–271.

Sauerland, Uli and Penka Stateva. 2007. Scalar vs. epistemic vagueness: evidence from approximators. *Proceedings of SALT17,* 228-245.

Sevi, Aldo. 1998. *A semantics for almost and barely.* MA thesis, Tel-Aviv University.

Simonyi, Zsigmond. 1881. *A magyar kötőszók, egyúttal a magyar összetett mondat elmélete. [The Theory of Connectives and the Complex Sentence in Hungarian.]*

Simonyi, Zsigmond. 1888. *A magyar határozók. [Adverbs in Hungarian.]*

Oravecz, Csaba, Váradi Tamás, Sass Bálint. 2014. The Hungarian Gigaword Corpus. In: *Proceedings of LREC 2014.*