A universal constraint on sensory lexicon, or when *hear* can mean 'see'?

Elena Maslova
Stanford University, University of Bielefeld

1. Introduction

This paper examines one specific construction for *HEAR* in Kolyma Yukaghir, whose peculiar behavior may have, as I will try to show, far-reaching theoretical and typological implications. Such subtleties are too often overlooked in broad-scale cross-linguistic studies, the research projects of St. Petersburg Typological Group being one of the very few exceptions. It is a great pleasure to offer this paper to Viktor Samojlovich Khrakovsky, one of those from whom I have learned early enough in my linguistic life just how important it is to enhance a broad typological perspective with attention to minor details of each language's structure – and vice versa.

2. Constraint on *SEE-HEAR* polysemy and "PERCEIVE-languages"

The typology of intrafield polysemy patterns of sensory verbs (*SEE, HEAR, TOUCH, SMELL, TASTE*) was first explored by Viberg (1984); this study is based on a 53 language sample and reveals the following cross-linguistic distribution of polysemy patterns (here *FEEL* stands for any pattern that covers *TOUCH*, *SMELL*, and *TASTE*):

(1)　*FEEL* [vs. *SEE* vs. *HEAR*]　　　　　　　14
　　　HEAR + FEEL [vs. *SEE*]　　　　　　　　7
　　　SEE + FEEL [vs. *HEAR*]　　　　　　　　1　　　(Kurdish)
　　　SEE + TOUCH vs. *HEAR + SMELL + TASTE*　1　　　(Swahili)
　　　SEE + HEAR + FEEL　　　　　　　　　　1　　　(Kobon)

It can be easily observed that polysemy patterns that subsume visual experience (*SEE*) are least frequent (three languages), i.e., this sensory modality is most likely to be signified by a separate lexical item; *HEAR* is the next according to this parameter (eight languages), and other sensory experiences are expressed by the same lexical item much more frequently (twenty three languages). Most importantly for the further discussion, *SEE* and *HEAR* are signified by the same verb in one language only (Kobon, a Papuan language of New Guinea), where this verb (*nöN-* 'perceive; understand') also covers all other sensory experiences.

These findings suggest that there is a strong cross-linguistic preference for lexicalization of *SEE* and *HEAR* as distinct concepts. For the following discussion, it will be convenient to define two language types corresponding to two possible states of sensory lexicon; I will refer to languages that have distinct primary expressions for *SEE* and *HEAR* as **SEE-languages**, and to languages that have the same primary verb for both sensory experiences, as **PERCEIVE-languages;** note that here and below REGULAR SMALL CAPITALS are used for language types and *ITALIC SMALL CAPITALS* for verb meanings. Viberg's study shows that SEE type is significantly more frequent than PERCEIVE type among the world's languages; assuming that his sample is more or less representative, approximately 2% of all existing languages are in PERCEIVE state at the present time. To be more accurate, Australia is not represented in Viberg's sample, and this may somewhat "magnify" this tendency (see below), but its statistical significance in the current language population is beyond any doubt in any event.

The first question raised by such a strong cross-linguistic tendency is whether it can, at least theoretically, be due to accidental events in the history of the language population and thus totally irrelevant for linguistics. The "null hypothesis" would be that SEE and PERCEIVE states of sensory lexicon do not in fact differ in terms of "linguistic preferences" of any kind (processing, cognitive, etc.), which implies that these states have the same degree of diachronic stability. In other words, it is equally likely for a SEE-language to change to PERCEIVE state and for a PERCEIVE-language to change to SEE state.[1] This hypothesis can be easily rejected on statistical grounds. Indeed, even in the worst possible case, that is, if, for some pre-historical reasons, **all** languages of the earth were in the SEE-state at some moment of time, and assuming in addition that the change of sensory lexicon is a very slow process (e.g., it occurs, on average, for 1% of all languages per 1000 years), it would take only 10000 years to increase the frequency

of PERCEIVE-languages in the population to 10%. After 50000 years (the **minimal** time of existence of the language population according to most received estimates), there would be no less than 30% of PERCEIVE-languages at the present time even if the Proto-population consisted solely of SEE-languages. Thus, the current distribution allows to infer that the probability of PERCEIVE-to-SEE shift must have been generally higher than that of SEE-to-PERCEIVE shift in the history of language population, or to put it the other way round, PERCEIVE type has been less diachronically stable. This means that this distribution reveals a linguistically significant phenomenon (rather than being an accidental byproduct of human history), namely, a universal constraint on "optimal sensory lexicon" that strongly affects the processes of lexical change. Following Optimality Theory (OT) notational conventions, I will refer to this constraint as *SEE = HEAR*:

(2) *SEE = HEAR*: Avoid referring to visual and auditory experiences by expressions that are ambiguous between *SEE* and *HEAR*.

A less trivial question is whether SEE-to-PERCEIVE shifts are possible at all. Indeed, it may be the case that the modern PERCEIVE-languages (like Kobon) retain the structure of sensory lexicon characteristic of a previous stage of linguistic evolution (e.g., of the Proto-language(s)), which cannot be acquired by a language again after it achieves SEE state (**strong unidirectionality hypothesis**). This hypothesis implies that *SEE = HEAR* actually works as a "hard" constraint at the present time, yet "the present time" in this formulation is just too short an interval for the lexical change triggered by this constraint to have taken place in all languages of the population (so far). In other words, PERCEIVE-languages exist because *SEE = HEAR* is a (relatively) recent phenomenon in the history of language population and its impact on the overall cross-linguistic distribution is still incomplete. The alternative hypothesis (**soft constraint hypothesis**) is that *SEE = HEAR* works as a "soft" violable constraint (which can be, in principle, overridden by other constraint(s) on the structure of sensory lexicon). This implies that SEE-to-PERCEIVE shifts are possible (but just significantly less frequent than the reverse shifts). Statistically, both hypotheses are quite plausible: under the strong unidirectionality hypothesis, a distribution very close to the actual one would be observed, for instance, if all languages were in PERCEIVE state 50000 years ago and the average frequency of PERCEIVE-to-SEE shifts during this period was approximately 8% of PERCEIVE-languages per 1000 years (under the assumption that the reverse change is impossible). Alternatively, if SEE-to-PERCEIVE shift is possible but, say, fifty times less probable than the reverse one, then the type-shift process is also bound to bring about a similar distribution in the long run (the time interval needed to achieve this distribution depends on the overall rate of the process and is irrelevant for the present purposes).

Thus, the attested cross-linguistic tendency can be accounted for in terms of two equally statistically plausible hypotheses, which have entirely different theoretical and empirical implications. The strong unidirectionality hypothesis entails that the existence of PERCEIVE type reveals an essential distinction between two stages of language evolution (PERCEIVE stage vs. SEE stage) and predicts that, in the long run, all languages in the world will shift to SEE state. Under the soft constraint hypothesis, the current distribution gives no clues for reconstruction of language evolution, but strongly suggests that there exist certain processes of "lexicon optimization" governed by OT-style violable constraints. In this case, *SEE = HEAR* makes PERCEIVE state diachronically unstable (i.e., if a language gets into this state, it is likely to create separate expressions for *SEE* and *HEAR* within a relatively short period of time), but achievable (i.e., a language can occasionally also change from SEE type to PERCEIVE type). If so, then a small percentage of PERCEIVE-languages (as now) is a relatively stable property of the language population. Given that statistics can neither reject nor confirm either of the contrasting hypotheses, the crucial linguistic question to be answered is whether or not SEE-to-PERCEIVE shifts are in fact possible, in particular, whether a verb for *SEE* can be extended to cover *HEAR* and/or vice versa.

3. *SEE > HEAR*: semantic extension or lexical specification

Viberg (1984) directly addresses the issue of diachronic changes in the structure of sensory lexicon; his major claim with regard to two primary sensory modalities is that a verb for *SEE* can undergo **semantic extension** to subsume *HEAR*, but not vice versa (**weak unidirectionality hypothesis**). Viberg's study has been recently extended to Australian languages (missing from his sample) by Evans and Wilkins (2000), who support this claim and contend that constraints on intrafield extensions in sensory lexicons are determined by "neurophysiological givens (the structure and experience of basic perception)" – in contrast

with transfield projections, which are open to cultural variation (2000:547).

The weak unidirectionality hypothesis consists of two distinct claims, which must be discussed separately:

(3)   Weak Unidirectionality Hypothesis:
        (a) The semantic extension *SEE > HEAR*, and hence the type shift SEE-to-PERCEIVE, are possible.
        (b) The semantic extension *HEAR > SEE* is impossible.

From the vantage point of the problem stated in the previous section, the most important part of weak unidirectionality hypothesis is (3a): if it is correct, then the problem is solved in favor of the soft constraint hypothesis. Examples invoked in (Viberg 1984, Evans and Wilkins 2000) to support (3a) fall into two groups. In one group of languages, the basic expression for *HEAR* is a combination of *SEE* with something else (X):

(4)   *SEE*   *HEAR* = X + *SEE*
Lesgian (East Caucasian)   *akun*   **van** *akun*   Viberg 1984:140
Warlpiri (Australia)   *nyanyi*   **purda** *nyanyi*   Evans and Wilkins 2000:557
Diabugay (Australia)   *ngundal*   **bina** *ngundal*   ibid.

The second group are PERCEIVE-languages that are reconstructed to have developed from SEE-languages by way of *SEE > HEAR* extension (all these languages are Australian):

(5)   Yir Yoront   *Karr*   'see, look at; hear, listen'   Evans and Wilkins 2000:556
        Gugu Yalanji   *Nyajil*   'see, hear, perceive'   ibid.
        Guugu Yimidhirr   *nhaamaa*   'see, look at, hear; think'   ibid.

Finally, Kobon (which is not considered by Viberg as an instance of *SEE > HEAR* extension) seems to be in an intermediate state: the auditory experience can be signified not only by the verb *nöN-* 'perceive, understand', but also by its combination with a particle (which does not occur in any other contexts), *apdi nö⚤-* 'listen to, hear', whereas *SEE* is always expressed by *nöN-* alone (Viberg 1984:150-151).

It seems, however, that the lexicalization pattern exemplified in (4) is at least as likely to arise as the result of **lexical specification** of a verb for *PERCEIVE*, i.e., in the course of PERCEIVE-to-SEE shift. Indeed, assuming universal *SEE = HEAR constraint (see Section 2), the speakers of a PERCEIVE-language would tend to disambiguate expressions for sensory experiences, and a natural way to do so is to add something (X) to the verb for *PERCEIVE* in some contexts; this situation is apparently attested in Kobon. Once this additional component becomes obligatory, the shift to SEE state is complete, and the language displays a pattern like in (4). The fact that in all attested examples of formal overlap it is *HEAR* (and not *SEE*) that receives overt marking indeed suggests that visual perception always (or at least more frequently) constitutes the **prototype** of "basic perception" in human languages, i.e., if a language has a single verb for *SEE* and *HEAR*, its semantic structure is organized around *SEE* as the prototypical meaning.[2] If so, then this semantic structure may indeed be determined, in Evans and Wilkins' wording, by "neurophysiological givens".[3] Yet the notion of prototype as a part of verb's semantic structure is fully compatible with both diachronic scenarios: under the scenario of lexical specification, a verb for *PERCEIVE* = *SEE* + *HEAR* with *SEE* as its semantic prototype naturally gives rise to a pattern like in (4); on the other hand, the meaning of a verb for *SEE* (as opposed to a different verb (Y) for *HEAR*) can be extended to cover *HEAR* and replace Y in this meaning. Note, however, that if the latter process involves an additional linguistic item (as in (4)), then it would not result in SEE-to-PERCEIVE shift: a language will still have a separate verb for *HEAR*, just a different one.

The problem is that neither Viberg nor Evans and Wilkins consider the specification scenario as a possible diachronic hypothesis, the only issue addressed being the direction of diachronic extension. One reason for this may be a widely received intuitive concept of typological plausibility of historical reconstruction, which implies that, other things being equal, it is "better" to reconstruct a frequent language type than a rare (let alone unattested) one. Indeed, if a randomly selected language is more likely to be in SEE state, then, as it seems, the same must be true for any reconstructed language as well. However

plausible this intuition might appear, it inevitably leads to self-contradictory results: if all (or almost all) reconstructed languages are in SEE state, then all (or almost all) reconstructed type shifts are SEE-to-PERCEIVE. Yet if such shifts were indeed more frequent than the reverse shifts, then the type-shift processes would have brought about a synchronic distribution in which there would be **more** PERCEIVE-languages than SEE-languages, contrary to the typological evidence that justifies the "more plausible" reconstruction in the first place. For example, Evans and Wilkins rely on reconstruction of a Proto-Australian verb **na-* for *SEE*. The theoretical possibility that this verb used to mean *PERCEIVE* (=SEE+HEAR) is not entertained, presumably because this verb is reflected as *SEE* (*LOOK AT*) in many Australian languages and as *PERCEIVE* only in a small subset of these languages (see two last examples in (5)). Yet these facts are fully compatible with specification (PERCEIVE-to-SEE) scenario as well, especially taking into account the attested cross-linguistic distribution, which indicates that PERCEIVE state must be diachronically unstable. If Proto-Australian happened to be a PERCEIVE-language, many of its descendants would have changed to SEE state by now, but some of them would retain the ancestor type (as in (5)). The apparent strong areal patterning of PERCEIVE-languages in Australia would be then accounted for in terms of a **single** cross-linguistically rare event (the PERCEIVE state of Proto-Australian), plus the universal preference for SEE state. On the contrary, Evans and Wilkins' scenario implies that Proto-Australian was a SEE-language, but quite a few of its descendants (seven out of sixty languages examined in the paper, 556-558) changed their sensory lexicon towards generally dispreferred state. In other words, this hypothesis assumes **multiple** cross-linguistically rare events, which for some mysterious reason occurred in the same area. Besides, under this hypothesis, most of Australian languages (not only PERCEIVE-languages) must have lost the original verb for *HEAR*: such a verb must have existed if Proto-Australian was in SEE state, but Evans and Wilkins mention no Proto-Australian reconstruction for *HEAR* (only for *EAR*).

To sum up, the specification (PERCEIVE-to-SEE) scenario appears to give a more typologically plausible account for the hypothesized instances of *SEE > HEAR* extension. This does not mean, of course, that this scenario is the "correct" one; this does entail, however, that, in absence of additional evidence to the contrary, the examples discussed here cannot be taken as compelling instances of SEE-to-PERCEIVE shift, and the question outlined in Section 2 remains unanswered.

4. A case study: *HEAR > SEE* extension in progress

In this section, I will present data of Kolyma Yukaghir[4] that appear to reveal an ongoing process of diachronic extension from *HEAR* to *PERCEIVE* (including *SEE*). This language thus provides a clear counterexample to the (weak) unidirectionality hypothesis put forward by Viberg (1984) and supported by Evans and Wilkins (2000). Besides, these data can be taken as a piece of evidence for the possibility of SEE-to-PERCEIVE shift, i.e., against the strong version of unidirectionality hypothesis (see Section 2).

Table 1. Kolyma Yukaghir perception verbs vs. Viberg's typological grid

Base selection:	Experiencer-based		Stimulus-based	
	Activity	Experience	Experience	Copulative
	'look at'	*'see'*	*'be/become visible, appear'*	*'look (e.g., tired)'*
Sense:				
sight:	*juö-de-*	*juö-*	**jed-, medu-**	***
hearing:	*mež-že-*	*medi:-, med-ej-*	*medu-*	***
touch:		*med-ej-*		
taste:		*med-ej-*		
smell:		*med-ej*		

Table 1 locates basic perception verbs of Kolyma Yukaghir within a typologically-oriented grid proposed by Viberg (1984). This grid is based on two parameters: the sensory modality signified by the verb (by means of eyes, ears, skin/body, tongue, and nose), and the type of situation construal: activities ('look at', 'listen'), experiences ('see', 'hear') and source-based copulative expressions ('look + complement', 'sound + complement'). In order to classify the Yukaghir perception constructions more accurately, it is necessary to draw a distinction between two semantic roles, **Stimulus** (an entity or event being perceived) and **Source** (the primary participant of an event being perceived). Generally, the Yukaghir perception verbs cannot take the Source NP as a core participant; for instances, it is impossible to

say that someone *heard the birds*, but only that they *heard the birds' song*. In what follows, I will denote an event being perceived (Stimulus) as E^S, the primary participant of this event as P^S, and the of perception as E^R.

As shown by the table, the Yukaghir languages have no copulative constructions with perception verbs (like *(to) look tired*). There are, however, two constructions[5] that do not straightforwardly fit into Viberg's scheme, but would apparently fall in the source-based class of expressions (in his terms), a **Stimulus-Based Experience Construction** with verbs ***jed-*** 'be/become visible' or ***medu-*** 'be/become audible, perceivable' (and their aspectual derivatives) and a **Source-Based Auditory Construction** (with verb *medu-* and its aspectual derivatives). The former construction is an instance of the basic Intransitive Construction: the Stimulus is the sole core participant. The Experiencer remains implicit; it may be either the speaker (Ex. (6)) or the main protagonist of a story (Ex. (7)). An experience is thus construed as a state (or change of state) of the Stimulus (rather than of the Experiencer), e.g.:[6]

(6)   *jel'o:d'e*   *okno-ge*   ***jed-*** *i*
      sun   window-LOC   visible-INTR(3SG)
      (lit.) 'The sun was visible in the window.' (= 'The sun was shining, thereby making itself visible (to me)').

(7)   *ta:t*   *shoromo*   *azhu:-k*   ***met-*** *t'i-l*
      then   person   word-FOC   heard-DLMT-SF
      (lit.) 'Then a person's word became audible (to her).' (= 'Someone said something, thereby making her hear their words').

The structure of the Source-Based Auditory Construction has no analogues in the Yukaghir inventory of grammatical constructions, but resembles Subject-to-Subject raising constructions in other languages: *medu-* serves the main predicate, the Source is the sole core participant, and the verb signifying E^S takes the form of Instrumental Action Nominal. The literal meaning of this construction is (approximately) 'P^S makes him/herself heard by way of E^R', e.g.:

(8)   *kel-delle*   *tude*   *touke*   ***an'n'e-l-e***   ***medi-s'.***
      come-SS:PFV   3SG   dog   speak-ANR-INST   be.heard-INTR(3)
      'He returned (home) and heard his dog barking.' [lit. 'His dog was heard
      by way of barking.']

The Source-Based Experience Construction overrides the lexical selection constraint on the verb *medu-* mentioned above, namely, the requirement that its subject refer to a sound or at least to an event associated with a sound (i.e., an appropriate Stimulus of hearing experience). In this construction, entities (e.g., a dog) can also "be audible", although their audible activity has to be specified by means of the Instrumental Nominal. The choice between the Stimulus-Based Experience Construction and the Source-Based Auditory Construction for hearing experiences depends on whether or not the Source (as a specific individual) is identifiable (from the Experiencer's point of view). In (7), the Experiencer hears someone speaking, but cannot recognize who is speaking; in (8), the Experiencer knows that he hears his own dog's barking (since he approaches his own dwelling).

The Source-Based Auditory Construction can also be used to describe **visual** experiences. Before we turn to the relevant examples, however, it is necessary to introduce one of the Experiencer-Based Experience Constructions. Kolyma Yukaghir consistently distinguishes two types of discourse contexts in which an observer is reported to perceive (*SEE* or *HEAR*) an event (E^S), depending on whether or not the listener is assumed to possess the information about E^S at the time of speech. If E^S has been previously mentioned, the event of perception (E^P) is described by means of the basic Transitive Construction with verbs *juö-* 'see' or *medi:-* 'hear' and E^S in the object slot; in this case, E^S is usually referred to by means of an anaphoric pronoun, but can also be expressed by a nominalized clause. Yet an event of perception is more often reported in order to **introduce** information about a new event (*And then he saw that…*). This discourse strategy is based on Event-Introducing Experience Construction, which formally resembles clause chains, with E^S referred to by the finite clause (in accordance with its intended discourse role) and E^P by a Different-Subject clause with an Experiencer-Based verb of perception. The Event-Introducing

Experience Construction differs from the regular Clause Chaining Construction in that the object (Stimulus) slot in the Different-Subject clause is absent, and the finite clause event must be interpreted as the Stimulus; that is, the finite clause in effect replaces the object slot in the valence pattern of Experiencer-Based verb. The following examples illustrate this construction for two primary sensory modalities:

(9)   ***juö-ŋi-de-ge***   *oqonastie*   *pulut*   *abuda:-l'el*
        see-PL-3-DS   Afanasiy   old.man   lie-INFR(3SG)
        'They saw that Afanasiy the old man was sleeping (there).'

(10)   *ta:t*   *irki-d'e*   *mikolaj-die*   ***medi:-de-ge***   *loci-n*
        then   once   Nick-DIM   hear-3-DS   fire-AT
        shanaha-k   ***medu:-l***
        crackle-FOC   be.heard-SF
        'Then once Nick heard fire crackling.'

Note that in (10) the event of perception is referred to twice, by the Different-Subject form of the Experiencer-based verb and by the finite Stimulus-based verb. As will be shown shortly, this "double marking" of E^P opens the possibility to encode two different aspects of experience within one construction.
       In the context of Event-Introducing Experience Construction The Source-Based Auditory Construction is commonly used to describe visual experiences, the actual sensory modality (i.e., *SEE*) being specified by the Different-Subject clause, e.g.:

(11)   *kebej-delle*   ***juö-ŋi-de-ge***   *alandin*   *tude*   *mahil-gele*
        go:PFV-SS:PFV   see-PL-3-DS   A   his   coat-ACC
        el-uldos'e-t   ***kel-le***   ***medi-s'***
        NEG-hang-SS:IPFV   come-ANR:INST   be.heard-INTR(3)
        'They went there and **saw** that Alandin was coming closer, his coat not hanging.'

(12)   ***juö-de-ge***   *okno-ge*   *pon'qo*   *piede-l-le*   ***medi-s'***
        see-3-DS   window-LOC   light   burn-ANR-INST   be.heard-INTR(3)
        'He **saw** that there was light in the window (lit. light burning).'

The question arises, Why E^S in these and similar examples is described by means of the Source-Based Auditory Construction, given that the event of perception is expressed in the Different-Subject clause, and in an apparent contradiction with the visual modality of the experience? For instance, why was Alandin in (11), literally, "heard by way of coming closer", rather than simply "coming closer"? As it seems, the function of this construction is to signal that the Source is being perceived (in this case, seen) by virtue of their own action(s). Compare examples (9) and (11): in (9), the protagonists saw the old man not because he was sleeping or, for that matter, not because he was doing something, but because **they** came closer to his campsite. In contrast with this, Alandin became visible to his associates because **he** was coming closer, that is, in a sense, he **initiated** the event of perception. More generally, the construction signals that the Experiencer's attention is attracted to the Source because of the Stimulus (E^S) it produces.
       This observation suggests that Kolyma Yukaghir categorizes primary sensory experiences according to two interrelated parameters: **sensory modality** (*SEE* vs. *HEAR*) and **initiator** (Experiencer vs. Source). The link between these parameters is motivated by the fact that a prototypical hearing experience is initiated by the Source (an entity must emit a sound in order to be heard, but can be seen without initiating any events). Since a visual experience also can be initiated by the Source, the Source-Based Auditory Construction is extended to subsume such experiences. It seems plausible to hypothesize that this process has been facilitated by the context of Event-Introducing Experience Construction, which allows the speaker to specify the "right" sensory modality in a different slot. It appears, however, that the process of *HEAR* > *SEE* extension is beginning to affect other Auditory constructions as well, e.g.:

(13)   *medin*   ***me-medi:***   *qojl*   *mon-d'e*   *shoromo*
        just   AFF-hear(1SG)   God   [say-REL]   person
        'It is the first time that I **see** someone who prays God.'

As mentioned above, the verb *medi:-* in its prototypical meaning requires that the object slot be occupied by the Stimulus; thus, the Source NP in the object slot in (13) can be considered as a piece of evidence for the semantic shift reflected in the translation (the idiomatic translation is based on the original Russian gloss). On the other hand, the sentence obviously implies that it is the Source's action (praying) that attracts the Experiencer's attention, i.e., this usage of the transitive verb *medi:-* appears to occur under the semantic conditions established above for the Source-Based Construction. It can be argued that this example is unconvincing, since the speaker apparently both sees the praying man and hears him pray at the same time; however, this is precisely the point: in the situation where both Experience-Based verbs would seem appropriate, the speaker chose *medi-* 'hear' and not *juö-* 'see', i.e., HEAR is extended to cover a visual experience and not vice versa.

To sum up, I hope to have shown that Kolyma Yukaghir is undergoing the process of HEAR > SEE extension, which has started with the Source-Based Auditory Construction, but has inevitable repercussions for the lexical meaning(s) associated with the verb stem for HEAR (*med-*), which are occasionally manifested in other constructions as well. Although it would be hardly plausible to "predict" with any degree of certainty that this process will result in SEE-to-PERCEIVE shift (see Section 2), i.e., that *med-* will eventually replace *juö-* as the basic verb of visual perception, yet this case study shows the theoretical possibility of intrafield semantic extension which may eventually lead to such type shift. It should be stressed that the **semantic** motivation of the attested extension – that is, the fact that an auditory experience implies some activity of the Source, while a visual experience can but need not be associated with the Source's activity – has nothing culturally specific to it; rather, to repeat Evans and Wilkins formulation once more, it is determined by "the structure and experience of basic perception" (547). Therefore, the attested reinterpretation of 'HEAR P' as 'PERCEIVE P BY VIRTUE OF P'S ACTIVITY' and the resulting extension to 'SEE P BY VIRTUE OF P'S ACTIVITY' is hardly determined by the Yukaghir cultural environment and thus can be thought of as an open possibility for any language.

5. Conclusions

Kolyma Yukaghir data show that a combination of language-specific conditions may open the possibility of semantically (cognitively) feasible extension of an "auditory" construction towards a broader meaning that subsumes visual experiences. Although this is only the initial step on a path that may lead to eventual SEE-to-PERCEIVE shift, yet this is the nature of shifts in progress: they cannot be complete by definition. Therefore, I believe that this case study gives an important piece of evidence for the hypothesis of *SEE = HEAR as a violable ("soft") constraint and thus against the strong version of Unidirectionality Hypothesis. *SEE = HEAR thus emerges as an example of soft (but rather powerful) universal constraint that significantly affects **stochastic** regularities of lexical change and thus determines a strongly skewed synchronic cross-linguistic distribution of languages between PERCEIVE and SEE types which nonetheless shows a small percentage of PERCEIVE languages. This entails that PERCEIVE type cannot be considered as a trace of a previous stage of language evolution; it is a rare but attainable typological option.

On the other hand, Kolyma Yukaghir obviously falsifies Viberg's (1984) weak unidirectionality hypothesis, since it demonstrates that a verb for HEAR can be extended to cover visual experiences. It does not undermine, however, the **synchronic** version of this hypothesis tentatively proposed in Section 3, which is repeated here: if a language has a single basic verb for HEAR and SEE, that SEE is its prototypical meaning. Since Kolyma Yukaghir has a distinct basic verb for SEE, it cannot falsify this hypothesis; it may be the case that if and when the verb for HEAR takes over the domain of visual experience completely, its (synchronic) semantic prototype will shift to SEE due to universal properties of human perception.

References

Croft, William. 2001. *Radical Construction Grammar. Syntactic Theory in Typological Perspective.* Oxford: Oxford University Press.

Evans, Nicholas, and David Wilkins. 2000. In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language* 76-3: 546-592.

Greenberg, Joseph H. 1995. The diachronic typological approach to language. In: Masayoshi Shibatani and Theodora Bynon (eds.) *Approaches to Language Typology*, 143-166. Oxford: Clarendon Press.

Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 2000/3: 1-62.

Viberg, Ake. 1984. The verbs of perception: a typological study. In: Butterworth, Brian, Bernard Comrie and Östen Dahl (eds.) *Explanations for language universals*, 123-162. Berlin: Mouton.

---

1. See (Greenberg 1995) for a discussion of diachronic approach to typology and (Maslova 2000) for a detailed explication of the relation between transition probabilities and synchronic statistical tendencies.

2. It should be noted, however, that there is no conclusive evidence against the possibility of the opposite marking pattern (*SEE* = *HEAR* + X); such examples may just remain undiscovered so far.

3. Indeed, we can hypothesize that, say, in a dog language, *HEAR* might probably be expressed as *EAR-SMELL* or the like, due to their different "neurophysiological givens".

4. Kolyma Yukaghir is spoken in Northern Siberia and belongs to a small isolate language group, which includes only one other language, Tundra Yukaghir. The process described here is not attested for Tundra Yukaghir, although these languages have very similar sensory lexicons.

5. As will be clear from the contents of this section, the relevant Kolyma Yukaghir data requires consistent references to perception **constructions** (rather than to perception **verbs**), which calls for Radical Construction Grammar style of description (Croft 2001). I adopt the common orthographic distinction between Capitalized Names of specific Kolyma Yukaghir constructions and general terms for cross-linguistic construction types.

6. All Kolyma Yukaghir examples come from authentic texts. Abbreviations: ACC – Accusative, AFF – Affirmative, ANR – Action Nominalizer, AT – Attributive, DIM – Diminutive, DLMT – Delimitative, DS – different subject, FOC – focus, INFR – Inferential, INST – Instrumental, INTR – Intransitive, IPFV – Imperfective, LOC – locative, NEG – Negative, PFV – Perfective, REL – Relative (clause), SF – Subject-Focus (verb form), SG – Singular, SS – same subject.