The 3rd Usage-Based Linguistics Conference
3-5 July 2017, Jerusalem
Session 1: Language Change (Chair: Mira Ariel)
Hebrew *hevi*’s path towards ‘give’: usage-based all the way

Roey Gafter (Ben Gurion University of the Negev), Scott Spicer (Fulbright Postdoctoral Fellow, Northwestern University) & Mira Ariel (Tel Aviv University)

In standard Hebrew, *hevi*, usually glossed as ‘bring’, contrasts with *natan*, ‘give’. However, while this may suggest that Hebrew, like English, clearly distinguishes giving events from bringing events in its verbal semantics, the usage patterns of these verbs are in fact more complex. Kuzar (1992) claimed that in recent years there is an expansion of the meaning of the verb *hevi* towards that of the verb *natan* – that *hevi* is gaining the meaning of ‘give’ alongside ‘bring’. Although examples such as (1) seem to point to a renewal of the ‘give’ meaning by recruitment of an innovative form (*hevi*), we demonstrate that such a “renewal” is not the motivating force for the change. We thus support Reinöhl and Himmelmann (to appear), who subsume traditional renewal cases under a general process of grammaticization. Ours is a case of semanticization.

We examine a corpus of Hebrew blogs (Linzen 2009), and demonstrate that there is indeed an ongoing change in progress in the meaning of *hevi*. The results show a significant effect of the age of the speaker (p<0.001) – older speakers are more likely to use *hevi* for unambiguous *BRING* events, whereas younger speakers are more likely to use it in contexts which are also compatible with giving events, as in (2). However, contra previous claims, we argue that *hevi* has not (yet) acquired the full range of ‘give’ meanings associated with *natan*. Crucially, both our corpus data and a follow up acceptability judgment study show that most speakers do not use *hevi* for ‘give’ when the event cannot (at least in principle) also be construed as a bringing event: examples such as (3) are exceedingly rare in our corpus, and are disfavored in the acceptability judgment experiment.

We outline a step-by-step semanticization cline for this change (see Traugott and Dasher 2002), proposing a much more nuanced innovative use of *hevi*, where for the most part the event construed is compatible with both ‘bringing’ and ‘giving’ as in (2). We further argue that it is usage patterns, namely, the specific discourse profiles associated with *hevi*, that account for the change in meaning. To turn into ‘give’, *hevi* must lose its ‘agent motion’ component and add a ‘caused possession’ component. Indeed, we demonstrate (i) a frequent nonliteral use of *hevi*, as well as (ii) a frequent use of *hevi* within the double object construction. The former cases lack actual motion, which contributes to the bleaching of ‘agent motion’, and the latter cases add on a ‘caused possession’ interpretation, which was originally contributed by the construction (Goldberg 1995), but is eventually lexicalized for the verb.

A comparison with English corpus data (The Santa Barbara Corpus), where *bring* remains distinct from *give*, supports our analysis in that *bring*’s discourse profile is quite different from that of *hevi*: *Bring* is not used nonliterally as often as *hevi*, and it rarely occurs in the double object construction (even compared to the older speakers in our sample, who rarely use *hevi* as ‘give’).

We are thus able to show the potential causal connection between usage and meaning, both positively (in Hebrew) and negatively (in English). An independent pull towards renewing ‘give’ need not be assumed here. Rather, what seems to be a renewal is in fact the semanticization of frequent contextual inferences (‘lack of motion’, ‘caused possession’), which are currently being incorporated into an additional lexical meaning for *hevi*. 
Examples

(1) ani homles lama še-lo tavi l-i kcat kesef
I homeless why that=not (give) me little money
‘I’m homeless, why don’t you hevi (give) me some money?’

(2) hi hevi-a l-i be=matana ner lev xamud
She hevi.PST-3F to-me in=present candle heart cute
‘She hevi (brought/gave) me a cute heart candle as a present’

(3) ima hevi-a l-i dira be=matana
Mother hevi.PST-3F to-me apartment in=present
‘(My) mother hevi (gave) me an apartment as a present’

References

Chicago, IL: University of Chicago Press.


Linzen, Tal. 2009. Corpus of blog postings collected from the Israblog website. Tel Aviv: Tel Aviv University
Reinöhl, Uta and Nikolaus P. Himmelmann. To appear. ‘Renewal’: A figure of speech or a process sui
generis?. *Language*.

The road to auxiliariness: a usage-based diachronic account

According to a widely-cited statement by Bolinger, ‘the moment a verb is given an infinitival complement, that verb starts down the road of auxiliariness’ (1980: 297). However, to the best of our knowledge, there have been few attempts to explain how verbs acquire infinitival complements in the first place. The goal of this paper is to provide such an explanation for a single class of lexical verbs that are known to grammaticalize into tense-aspect auxiliaries, namely, verbs meaning FINISH, which is understood here as verbs that denote the act of bringing a process to completion.

In this talk, we present a detailed analysis of the grammaticalization of FINISH antecedents on the basis of diachronic data that aims at answering these questions. Spanish developed a FINISH anterior comprising the erstwhile lexical verb acabar ‘finish,’ followed by the preposition de ‘of,’ which in turns governs an infinitive (see 1).

(1) Juan acab-a de comprar un coche
   ‘Juan just bought a car’

We conduct a logistic regression analysis over of all 1962 tokens of FINISH constructions in a corpus of Spanish historiographical texts from the mid-13th to the end of the 18th century. Our main findings are two. First, we identify a hitherto undescribed process of language change in which previously inferred meanings are later obligatorily expressed in an overt fashion, a process which we call here overtification. In this case, acabar initially occurs without a verbal complement when the finished event is the default interpretation (e.g., finish the bridge -> finish [building] the bridge in (2)). In contrast, acabar de + infinitive occurs when the finished event differs from the default interpretation (e.g., finish rebuilding the city in (3)). In other words, the acabar de + infinitive construction signals that that there is something unexpected or particularly informative about the finished event.

(2) e acab-ó la torre dal=faro que
    and finish-PST.PFV.3SG DET.DEF.F.SG tower of.the=light.house that
    començ-a-ra hercules
    begin-theme-THEME-PST.IPV.SBJ.3SG Hercules
    ‘And he finished the tower of the light house than Hercules had begun’ (13th c.)

(3) e despues que este rey dario fue muer-to […]
    and after that DET.DEF.M.SG king Dario be.PST.PFV.3SG die-PTCP
    xerxes su fijo […] acab-ó de redificar la
    Xerxes his son acabó de rebuild DET.DEF.F.SG
    cibdad city
    ‘And when King Dario had died […] Xerxes his son […] finished rebuilding the city’ (15th c.)

Second, the temporalization of the construction - the emergence of its anterior function - followed its generalization to uninformative contexts, exemplified by examples in which a stereotypical event is overtly expressed in the acabar de + infinitive construction. In (4), from the 16th century, the acabar de + infinitive construction no longer codes informativeness. In other words, what was previously inferred was now explicitly expressed, the result of the process we have called here overtification.

(4) ...en este tiempo como fuese acab-a-da de
    in DET.DEF.M.SG time when be.PST.IPV.SBJ.3SG finish-THEME-PTCP.F
    hacer la puente pas-ó la Infantería española
    make DEF.DEF.F.SG bridge pass-PST.PFV.3SG DEF.DEF.F.SG infantry Spanish
    ‘And then, when the bridge had been finished to be built, the Spanish infantry passed over it’
In cases such as (4), writers use *acabar + de + infinitive* to highlight the fact that a narrated (finished) event is important to the progression of the discourse. Crucially, the possibility of this rhetorical use of the construction results from the original informativity-coding function of *acabar + de + infinitive*; if a narrated event is perceived as informative, it is likely to be relevant to discourse progression. The rhetorical use of *acabar + de + infinitive* caused the semantic reanalysis of the construction by the hearers. When speakers used *acabar + de + infinitive* in contexts that were not in conformance with its original function of marking informativity, hearers of the construction were not able to accommodate the presupposition that the proposition is highly informative (after all, the speaker could have used the simple *acabar* construction to express the same proposition), resulting in Pragmatic Overload (Eckardt 2009).

To avoid this pragmatic overload, the hearers had to reanalyze the meaning of the construction and took the context as the basis for this reanalysis. Because *acabar + de + infinitive* was frequently used in perfective past-of-past contexts, the hearers reanalyzed the construction as a marker of recent past. In line with this assumption, our analysis shows that such perfective past-of-past contexts served as catalysts for both the temporalization of *acabar + de + infinitive* after the 17th century (Fig. 1).

Fig. 1. Interaction effect between the variables Time, Subordinaton and Perfectivity in the logistic regression model

In summary, we propose that the change described in this talk is motivated by inferential principles rooted in online speaker-listener interaction. The emergence of verbal complements - the first step down Bolinger’s road to auxiliariness - is motivated by informativity: the use of an overt verbal complement of *acabar* was an explicit signal to interpret the finished event as informative with respect to the presupposed default interpretation. Once speakers used the construction in previously unlicensed contexts, listeners were coerced into an innovative form-function pairing. The innovative temporal meaning was not ‘out of the blue,’ but rather listeners took the typical usage context as a cue for its meaning.

References


Differential Recipient Marking in Romance languages: a diachronic typology and its cross-linguistic implications

This paper presents a usage-based account for the distribution and grammaticalization of differential marking patterns (also called differential “flagging”), focusing on previously undescribed patterns of differential Recipient marking (DRM) in Romance languages. In DRM, commonly recognized in languages of Asia and the Americas (Kittilä 2011), a subset of free (pro)nominial arguments associated with the semantic role of Recipient may or must differ in morpho-syntactic coding from other such arguments. Unlike differential marking of subjects and objects, DRM was not previously studied diachronically (cf. Malchukov 2008; Kittilä 2011).

Based on diachronic and web corpus studies in Spanish and French, I demonstrate that Romance languages attest to typologically opposite DRM patterns. Synchronically, these patterns differ morphologically and semantically: the Spanish pattern use a comitative-based differential marker (‘wiith’) for highly individuated arguments (1), while French varieties display an apudlocative-based (‘at, to’) to differentially mark unindividuated referents (Hopper & Thompson 1980). These marking patterns contrast with the expected dative-marking (>Latin ad).

I propose a diachronic account for this typological split. The differentiation in ARGUMENT MARKING (16th century-present), including DRM, was affected by the Old Romance split in ARGUMENT INDEXING patterns by bound morphemes (e.g. clitics, affixes)(11th-13th centuries). In other words, clitic and affix morphology grammaticalized differential coding centuries before similar patterns grammaticalized in free argument structure (e.g. by prepositional marking). This may be explained as due to the grammatical entrenchment of individuation-based indexing tendencies affecting the grammaticalization of typologically similar marking patterns. Divergent patterns (=indexing and marking display opposed tendencies) usually involve language contact.

Additionally, I propose two grammaticalization pathways for the patterns in (1) and (2) and illustrate how they contribute to the typology of argument structure. The proposed pathways share common properties with a recognized grammaticalization pathway of “regular” dative markers (Creissels 2009)(3), suggesting that similar inferences are at play in change to differential and non-differential markers. The pathway in (3) reads such that I Recipient markers in Romance first mark location in place (‘be at/with’), then motion towards Goal (‘go to’) and finally caused motion towards Recipient (‘give’, ‘send’). However, certain differences between the proposed pathways in Spanish and French have far-reaching implications for the diachronic typology of argument marking. In contrast with previous research, they illustrate that:

(ii) An implicational hierarchy of the grammaticalization of differential marking patterns across argument positions follows from observations on DRM and its relation to differential marking of subjects and direct objects.
(iii) When combined with insights from typology of argument structure, this implicational hierarchy causally relates the distribution of differential marking types with the emergence of new non-differential object markers.
Differential Recipient Marking in Romance languages: a diachronic typology and its cross-linguistic implications

1. Spanish
   Si se=vea (Caroline) pues, la=envío contigo
   if se(.RFL)=go.PRSSBJ.3SG then 3SGF.ACC=send.PRS.1SG COM-2SG
   ‘If she goes (Caroline), so I send her over to you.’ (literally: ‘with you’)(COM=comitative)

2. French
   J’=ai envoyé un mail chez le carrossier
   1SG.SBJ=have.PRS.1SG send.PTCP a mail APUD the vehicle-bodyworker
   ‘I sent a mail to the vehicle bodyworker.’ (literally: ‘at the vehicle bodyworker’)(APUD=apudlocative)

3. adessive (juxtaposition near reference-point) → apudlative (Goal-marking in intransitive clauses) → apudlative (Recipient-marking in ditransitive clauses).


The on-line emergence of syntactically unintegrated post-positioned she- (‘that/which/who’) clauses in casual spoken Hebrew talk

Yael Maschler (Haifa University)

So-called subordinate clauses which have loose or no syntactic relations to an element in a preceding clause remain relatively unexplored, but they have received some attention recently (e.g., Evans 2007; Mithun 2008; Laury and Seppänen 2008; Keeverlik 2008; Verstraete, D’Hertefelt, and Van Linden 2012; Mertzlufft and Wide 2013; Günthner 2014; Wide 2014; Evans and Watanabe 2016). Mithun explores this phenomenon from a wide typological perspective, showing that grammatical dependency markers can be functionally extended “from sentence-level syntax into larger discourse and pragmatic domains” (2008: 69) to mark “supplementary information that is not part of the storyline, material that sets the scene in narrative, contributes commentary or explanation, provides emotional evaluation, and so forth” (ibid.: 99).

To the best of my knowledge, such clauses have not been studied in spoken Hebrew discourse. According to traditional Modern Hebrew grammar, the monomorphic element she- (‘that/which/who’), attached to the following word, is known as the general, most frequent Hebrew ‘subordinator’ employed at the opening of the subordinate clause in bi-clausal syntactic constructions of the relative and complement type. The morpheme she- also attaches to additional components at the opening of a variety of adverbial clauses, forming various adverbial conjunctions: e.g., lifney she- (‘before’), kshe- (‘when/while’), mipney she- (‘because’), bemikre she- (‘if’). Finally, she- also opens the subordinate clause in a pseudo-cleft construction. However, investigations into the syntax of spoken Hebrew discourse “taking temporality (Hopper 1987, 2011) seriously”, show that this traditional description of Hebrew subordination is oftentimes lacking (Maschler 2011, 2012, Polak-Yitzkhaki and Maschler 2016, Maschler and Fishman forthcoming).

My paper examines she- constructions which challenge this traditional description because they are employed in a way that makes it unclear which, if any, subordinate category (relative, complement, adverbial, or pseudo-cleft) they constitute. Rather than attempt such a classification, I seek to explore the actions performed by these borderline she- tokens in interaction and their on-line (Auer 2009) emergence in interaction.

Examine, e.g., the following excerpt from a conversation in which three girlfriends are gossiping about a mutual acquaintance, Liron, criticizing her for choosing partners based on their socio-economic status:

'Liron's Boyfriends':

90 Moran: vekol sheni vexamishi, and every Monday and Thursday,
91 hi holexet le-recital,
92 she=!Liron! goes to-- Recital (=name of restaurant),
93 le'exol 'im 'eh.. 'ima ve'haxaver shel ha'imah she!Liron!,
to eat with uh.. mother his and the boyfriend hers
to eat with uh..his!Liron’s boyfriend’s! mother and her boyfriend,
94 ...shehaxaver shel ha'imah shelo,
95 ...she- the boyfriend of the mother his
96 ...she- his mother’s boyfriend,
97 ...hu profesor,
98 he professor
99 ...is a professor,
100 po.
101 here.
102 ...bekalkala.
103 In Economics.

The she- opening line 93 does not begin a normative relative clause, because the form correferential with what might be considered the antecedent, haxaver shel 'her [= Liron’s boyfriend’s mother’s] boyfriend’ (92), is a full NP that is actually much heavier than the ‘antecedent’ it supposedly refers back to: haxaver shel ha'imah shelo (lit. ‘the boyfriend of his [=Liron’s boyfriend’s] mother’, 93). Furthermore, the NP at 93 is altogether
‘out of place’ syntactically. Without it, she-lines 94-96 would constitute a normative relative clause - ‘she goes... to eat with his mother and her boyfriend, shehu profesor po bekalkala (‘who is a professor here in Economics’). Semantically, an analysis of lines 93-96 as a circumstantial adverbial clause expanding the main clause is also possible: ‘she goes ... with his mother and her boyfriend; his mother’s boyfriend being a professor here in Economics’. However, an adverbial clause modifies the predicate of the ‘main’ clause, not one of its nominal constituents, as here. Furthermore, the conjunction connecting such circumstantial clauses should be kshe- (‘while’), not she-. Thus we see that this she-clause is a borderline case, in between the relative and the adverbial categories, not quite fulfilling the formal criteria for either one. Furthermore, since there is no clear component within a ‘main clause’ which this she-clause substitutes, its subordinate status as embedded in a main clause is altogether questionable.

In this so-to-speak competition between girlfriends over who has the most ‘juicy’ gossip about Liron (from which the lines above are excerpted), the she-clause functions to convey the climactic piece of gossip as well as the speaker’s ‘outraged’ stance towards it. The marked prosody of line 93, along with the fact that ‘ima (‘mother’) is unusually preceded by the definite article, support this interpretation. Thus we see here the extension of the dependency marker she- to the discourse and pragmatic domains (cf. Mithun 2008). However, while in the languages studied by Mithun the function of such dependency markers is to denote background information, in this Hebrew example the dependency marker actually marks the climactic clause of the entire gossip session.

My paper explores a variety of syntactically unintegrated post-positioned she-tokens in a corpus of over 11 hours of audiotaped conversation, exploring the particular actions they implement in discourse and their on-line emergence in interaction, thus contributing to a pragmatic typology of syntactically unintegrated so-called subordinate clauses.

References


Maschler, Yael & Fishman, Stav. Forthcoming. From multi-clausality to discourse markerhood: The Hebrew ma she- (‘what that’) construction in so-called ‘pseudo-clefts’.


Session 2: Intersubjectivity (Chair: Yael Maschler)
The origins of a desiderative construction from indexing a first person perspective in reported speech

Linda Konnerth (Hebrew University of Jerusalem and University of Oregon)

We can easily express the intentions and desires that we ourselves have. This is because we (usually) know directly how we feel and what we plan. However, if we express the intentions and desires of somebody other than ourselves, we talk about knowledge that we do not have direct access to. This problem with expressing non-first person intentions and desires motivates the development of the main desiderative construction in Monsang (Trans-Himalayan/Sino-Tibetan), which reconnects non-first person desideratives back to a first person perspective via reported speech. This development is illustrated with data from a diverse corpus of spontaneous speech of 10,000+ words accumulated during extensive fieldwork on Monsang since 2014.

An example of a third person desiderative in Monsang is (1). The two components of the desiderative construction are (a) the desiderative suffix -mín on the lexical verb and (b) the inflected form of té, which we would synchronically analyze as an auxiliary.

(1) sí-mín  á-té-ná?
go-DESID  3SG-AUX:DESID-IPFV:TR
‘s/he wants to go’

However, comparison with a separate, semantically broader ‘reported speech/thought/intentionality’ (RSTI) construction in (2), makes it clear that á-té-ná? in the desiderative construction in (1) is etymologically ‘s/he says/thinks’, as is the case also in the RSTI in (2).

(2) [sí  kí-té]  á-té-ná?
go 1SG-AUX:SUGG:FUT  3SG-say-IPFV:TR
‘s/he says <I will go>’ / ‘s/he says that s/he will go’ /
‘s/he think <I will go>’ / ‘s/he thinks that s/he will go’ /
‘s/he wants to go’

This RSTI turns out to be functionally equivalent to what we can reconstruct as the source construction of the desiderative. It also consists of two components: (a) a verb of saying (á-té-ná? ‘s/he says’) plus (b) its complement, which is a first person future form (sí kí-té ‘I’ll go’). As for (a), the auxiliary part of the desiderative in (1) is identical with the verb of saying in the RSTI in (2) (i.e., á-té-ná?). As for (b), the lexical verb complex of the desiderative in (1) (i.e., sí-mín) requires an extra step of internal reconstruction in order to recognize that historically this first part of the construction, likewise, is equivalent to what we find in the RSTI in (2), i.e., a first person future form (sí ki-té).

Taking a closer look at sí kí-té ‘I’ll go’ or ‘let me go’ in the RSTI, there is an obvious question mark over the té glossed as the ‘suggestive future auxiliary’. Indeed, we find what must be grammaticalizations of té ‘say’ in three different future constructions in Monsang: in the ‘suggestive future’ as in (2); further in the common future sí-ván kí-té ‘I’ll go’; as well as in the inceptive future sí-rón kí-té-ná? ‘I’m about to go’. Grammaticalizations of reported speech to future constructions are also attested in Africa (Aaron 1996; Botne 1998) and in Papuan languages (Reesink 1993; Spronck 2016).

Underspecified RSTI constructions, such as the one in Monsang in (2), have previously been discussed for Australian languages by McGregor (1994; 2007) and Rumsey (2001), as well as recently by Spronck (2015). Similarly, desideratives and other constructions that express modality involve grammaticalizations of a verb ‘say’ in a number of languages around the world, including Trans-Himalayan ones (Saxena 1988; Noonan 2006). However, the case of Monsang, unlike anything found in Trans-Himalayan or elsewhere, clearly illustrates the grammaticalization of the RSTI to a dedicated desiderative, as well as an innovative RSTI construction. The corpus that the study is based on serves to identify the usage properties of the RSTI and...
therefore also those underlying the development into a dedicated desiderative construction. Finally, the future constructions apparently based on etymological té ‘say’ further speak to the diachronic versatility of reported speech constructions, as well as to their connections to modality more broadly.

References


Empathy and the wandering 1 and 2 person pronouns in Anal (Tibeto-Burman, Manipur)

An intriguing feature, characteristic of the South-Central (“Kuki-Chin”) group of the Trans-Himalayan/Tibeto-Burman language family is a high sensitivity to the involvement of Speech Act Participants (SAP) in an event. Expressed by interestingly different systems of pronouns and hierarchical verbal indexations, it reflects perplexing diachronic processes, attributed in the recent literature to socio-pragmatic challenges, involved in talking about “you and me”. For example, comparative studies reveal that the 1st person pronoun in one language may be the cognate of a 2nd person pronouns in a sister language, both owing their origin to an inclusive plural 1-person form (DeLancey 2014). Comparative studies of verbal indexation propose processes of avoidance, politeness and empathy, which result in complex indexation systems of SAP-related forms (DeLancey to appear). In particular, apparent 2-person indexes are found to take over other functions and mark SAP or even a 1-person P-argument (Konnerth 2017). However, so far there has been no empirical observation of any of the proposed social processes of person shift in the South-Central branch.

This preliminary study of a corpus of spontaneous natural speech in Anal, spoken by 20,000 speakers in the North-East Indian state of Manipur, sheds some light on the usage of 1st and 2nd person pronouns and some of these processes. In particular, it shows empathy-related usage of the 1st person dual pronoun (“you and me”) in place of 1SG or 2SG pronouns under relevant pragmatic conditions. The primary social strategy that triggers this usage can be seen as a representation of joint involvement in the event. This move demotes the responsibility of its actual sole agent. Consequently, 1DUAL in place of 2SG is a face-saving strategy in light of the addressee’s potential failure; while 1DUAL in place of 1SG appeals to the addressee’s involved attitude and evokes empathy. The former case can be seen in (1): although the action discussed is carried out exclusively by the addressee, it pursues the joint interest of the group. Consequently, it is expressed as if it was carried out jointly. This usage exhibits the speaker’s personal interest in a positive outcome, and the empathy for the addressee and his actions. It prompts downplaying the responsibility of the addressee, expressing that the action is performed for the sake of the group as a single whole.

This strategy paves the way for its symmetric usage by the speaker to diminish his personal responsibility. Such a usage involves even a larger gap between the representation and the actual event, as (2) shows. As a result, this bridging context triggers conventionalisation of 1DUAL as empathy-expressing 1SG marker. The usage expands also to events where no responsibility can be shared. Thus, in (3), the speaker appeals for the empathy of her significantly younger sister regarding her (i.e., the speaker’s) senior age. Similarly, in (4) the speaker uses a 1DUAL form, implying a regret about not visiting her native village for decades. However, the addressee has just returned from this village, as the continuation of the sentence demonstrates. In addition, (4) also shows that 1DUAL does not replace the 1SG: while the utterance evokes the pity of the listener through the dual form, it maintains the necessary 1SG-2SG opposition in the subsequent verb.

Finally, the 1DUAL prefix na-, which is gradually shifting to empathy evoking 1SG (example 5), is diachronically a 2SG prefix. It is preserved as a 2SG marker na- in nearly all of the surrounding languages, and is reconstructed as 2SG-marker for the proto-level of the branch. Thus, Anal exhibits a diachronic shift 2SG > 1DUAL > 1SG, motivated by social aspects of face-to-face interaction. Its final stage is currently ongoing and observable.
Examples:

The context for examples 1 and 2: a man attempts to demonstrate to a visitor traditional bow-and-arrow shooting. An observer who watches his attempts tells him (1). As he fails to pull the bowstring, he utters (2).

1. pə-< i >-tân pəchā:-sunŋ kã?-pʰó?-sin-te
< NMLZ >-aim be.able\1DUAL-SAP.COND shoot-ONCE\1DUAL-TOP
i-[hà-rʰaŋ] va
NMLZ-good-PURP COP

‘It will be good if you (lit: we both) will be able to aim and shoot it once.’

2. pətchà:-mà-sin-nú i-ṭìn
be.able-NEG\1DUAL-PAST VOC NMLZ-pull

‘I am (lit: we both are) not able, man, to pull it.’

3. sum-takʰ-a-t̪ər̪? i-vaŋ-va-sin-to
ten-seven-six NMLZ-COME-\1DUAL-TOP

‘Since I am (lit: we both are) getting 76 years old…’

4. ani-te pamː-te pamʰ-al-rʰō:-sin-nu
\1DUAL-TOP countryside-TOP forget-fully-\1DUAL-NF
ama-lu-e-kʰe ə:-pətɕːl:-nûm-páː-màŋ-nīŋ
3-like-that.NEAR.ADR-also 2-NON.AG-ask-want-very-PROG-1SG

‘Since I (lit: we both) completely forgot this village, I want to ask you how it is now.’

5. na-he-pədun-jaː-te na-paluŋ-kʰe veʔ-susu-ну
\1DUAL-DIST.PLAIN-think-JUST-TOP \1DUAL-heart-also feel.sorry-a.bit-PAST

‘As I (lit: we both) think about it, I (lit: our heart) feels a bit sorry.’

References:

Konnerth, Linda. 2017. 'Person', cislocative, and 'you (and me)': Recurrent pathways to SAP indexation in the South-Central branch of Trans-Himalayan. Talk given at Linguistics Seminar, Hebrew University of Jerusalem, January 10
Non-canonical hortative constructions in Hebrew and Russian: a contrastive analysis
Danny Kalev (Tel Aviv Univerity), Larisa Leisiö & Jiri Kuittinen (University of Eastern Finland)

Although Modern Hebrew (Hebrew henceforth) and Russian are genetically unrelated, a corpus-based analysis reveals that they both exhibit non-canonical hortative constructions that are based on past indicative forms. We will present these constructions and address the following questions:

- What are the motivations for the seemingly counter-intuitive choice of past tense forms?
- Are the non-canonical hortative constructions of Hebrew and Russian related?

Hebrew is a Semitic, tense-prominent language (Bhat 1999: 151). Its TAM categories are formed by means of interdigitation, i.e. interleaving vowels within the triconsonantal root. In Russian, an East Slavic aspect-prominent language (op. cit., 180), verbs are obligatorily classified as perfective or imperfective. Hebrew has two canonical imperative moods: the Inflectual Imperative and the Implicit Imperative (table 1). The latter is morphologically identical to the 2nd person future tense (Ariel 1990). Perfective verbs in Russian have inflectional past (table 2) and future forms. Imperfective verbs have present and past forms. The canonical hortatives in Russian include an imperative and a hortative construction. In both languages, past indicative forms may convey hortative and imperative meanings as well.

The cohortative mood indicates exhortation addressed to the speaker herself, either exclusively or including her addressees (Gesenius 1910: §108). Hebrew’s canonical cohortative consists of the auxiliary bošiłu ‘come.SG.M/SG.F/PL.’ +1st person plural future tense (example Error! Reference source not found.). The Russian canonical cohortative is based on the future indicative. In addition to these canonical constructions, in both languages past indicative forms (in Hebrew, 1st person specifically) convey a resultative cohortative that denotes an event viewed in retrospect – as if the speaker(s) are already in the future state ensuing from the event. The non-canonical cohortatives of Hebrew and Russian encode immediacy and often involve motion verbs (table 3). However, verbs from other semantic categories are permitted under certain restrictions: Hebrew requires dynamic verbs; Russian requires certain subtypes of perfective verbs (example 2).

The 2nd person past indicative forms in both languages may denote curt commands (examples 3–7). We term this construction the Military Imperative (MI). In Hebrew, this construction requires a phrase-initial temporal upper-bound (examples 6 and 7). We argue that unlike the canonical imperative constructions, the Hebrew MI is always complete, communicating commands that have to be executed “thoroughly and to completion” (Bybee et. al. 1994: 54). The Russian MI presupposes momentaneouness, completiveness or inchoativity. In both languages, the MI is considered harsh and therefore inappropriate in courteous exchanges – unlike the non-canonical cohortative.

With respect to the relative chronology, pošel, pošla etc. denoting curt commands are documented since the late 1700s (example 12). Hebrew’s non-canonical hortative constructions have surfaced only in recent decades. It seems plausible to suggest that these similarities are attributed to Hebrew-Russian contact. Indeed, the number of Russian speakers living in Israel is currently estimated at 1 million out of a total population of 8.6 million. That said, we suggest alternative explanations to these similarities. Recruiting past indicative forms to denote a resultative, i.e., an event viewed in retrospect, is motivated by iconicity. Similarly, motion verbs are universally recognized as a fertile source of TAM grammaticalization processes (Bybee et al 1994: 9).

To our best knowledge, Hebrew’s oldest documented non-canonical hortative is zaznu ‘Let’s move it!’ from 1951 (Keinan, 1952: 185). We assume that a subsequent expansion paved the way to a 2nd person past indicative hortative. Finally, the 1st person and 2nd person hortatives bifurcated, gradually transforming the latter into a harsh imperative. We suggest that this bifurcation occurred in both languages independently, due to similar pragmatic motivations: self-exhortation often involves a desirable or beneficial activity, whereas a request directed to another person is more likely to inconvenience the addressee, especially when immediate complete execution is expected. The structural differences between the MI forms in both languages seem to corroborate the bifurcation hypothesis. Therefore, we argue that contact may have served as a catalyst – rather than the trigger – in the above grammaticalization processes that eventually resulted in similar constructions in both languages.
Table 1: Hebrew Canonical Imperative forms of the verb ‘write’

<table>
<thead>
<tr>
<th>Future Tense, Implicit Imperative</th>
<th>Inflectional Imperative</th>
</tr>
</thead>
<tbody>
<tr>
<td>tixtov₂SG.M</td>
<td>kto₂SG.M</td>
</tr>
<tr>
<td>tixtevi₂SG.F</td>
<td>kitvi₂SG.F</td>
</tr>
<tr>
<td>tixtevu₂PL</td>
<td>kitvu₂PL</td>
</tr>
</tbody>
</table>

Table 2. Russian Indicative Past Perfective of *pojti* ‘go’ (inceptive, starting to go)

<table>
<thead>
<tr>
<th>Masculine</th>
<th>Neuter</th>
<th>Feminine</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>poš-el</td>
<td>poš-l-o</td>
<td>poš-l-a</td>
<td>poš-l-i</td>
</tr>
</tbody>
</table>

Table 3: Various Russian and Hebrew past indicative motion verbs with a cohortative construal

1. *bo* **nelex** ‘Let’s go’
   come.2SG.M.IMP go.1PL.FUT
2. *poter*-l-i **ruk-i** do tepl-a složi-l-i lados-k-i dom-ik-om
   rub-PST-PL.hand-PL until warm-GEN put-PST-PL palm-DIM-PL house-DIM-INS
   ‘Rub hands until they are warm, put your palms together like the roof of a house...’ (National Russian Corpus, Subcorpus of oral speech, Lecture on methods of sight restoration. 2004)
3. *ty* **začem** *zdes’?* poš-l-a! (Daškova 1786)
   2SG for.what here go-PST-F
   ‘[Lady to her maid] Why are you here? Go away!‘
   shut-PST-PL all mouth-PL
5. *svoboden* **soldat – skaza-l** “*ded*” – Uše-l ot televizor-a! (Šenderevič)¹
   footloose soldier – say-PST,M “old.man” – go.away-PST,M from TV-GEN
   ‘You footloose soldier!’ – the senior conscript said. “Move away from the television!”
   thirty seconds touch-PST.2PL at.DEF-fence return.PST.2PL
   ‘Be in a state of having touched the fence and returned to your position in 30 seconds!’
7. *shtey dakot* he’evartem et kol ha-mashash le-amud ha-xashmal sham
   two minutes transfer.2PL.PST ACC all DEF-gear to.pole DEF-electricity there
   ‘Be in a state of having moved (lit. you moved) all of the gear to the utility pole right there!’ [Rosenthal 2014]

¹ We are indebted to Alina Israeli and Olesja Khanina who drew our attention to examples 2 and 5, respectively.
References


Discourse Markers in Speech and Writing: Egyptian Arabic yaʕni (‘it means’) as Case Study

Differences between speech and writing are widely acknowledged to affect structural and functional variation across discourse types (see, e.g., Biber 1988; Chafe 1982). Discourse markers (DMs) provide a promising lens to examine these differences. As signals of socio-cognitive processes, relations, and actions that underlie discourse production and comprehension (see, e.g., Schiffrin 1987; Schourup 1985), DMs are highly instructive to our understanding of the various important ways in which spoken and written texts are distinct.

In this talk I explore the distribution and functions of the DM yaʕni (lit. ‘it means’) in spoken and written Egyptian (Cairene) Arabic. The analysis is based on data collected from a corpus of interviews which I conducted and recorded in Cairo in 2011, and a corpus of contemporary written prose. Egyptian Arabic provides a special case study for the examination of DMs in speech and writing, since it is a language with no long literary tradition, in which emergent conventions of writing – and their associated pragmatic and meta-pragmatic motivations – can still be discerned.

In a previous study, it has been observed that in unplanned speech yaʕni carries the basic function of signaling the interlocutor’s cognitive efforts to produce the locally most satisfying expression of his or her communicative intentions (Marmorstein 2016). This function underlies two distinct uses of yaʕni: (a) the introduction of focal (i.e., new and/or pragmatically unexpected) information, and (b) the framing of elaboration of some segment or aspect of prior talk. In both of its uses, yaʕni presents a symbiotic relationship (Östman 1982) with the socio-cognitive environment of natural spoken interaction, by facilitating and cueing the online processing and incremental verbalization of ideas, as well as the participants’ other-orientation.

For example, in excerpt (1), the speaker (H) initiates a story about her daughter (N), following a specific request to tell the story:

(1)  H:  

1. hiyya,  
   she  
2. N,  
   N  
3. yaʕni:,  
   yaʕni  
4. (o.6) galhaʕarīs.  
   came to her a bridegroom  
5. fi șahrîtalāta.  
   in month three

She, N, I mean a bridegroom came to [ask for] her [hand] in March.

H starts the story by referring to the already mentioned topic N (1-2). Then, to introduce the narrative, i.e., the new contribution made about this topic, she uses yaʕni. The clustering of yaʕni with silent (4) and filled pauses (3) evidences the increased cognitive labor involved in the processing of new or inactive information (cf., Chafe 1994), especially when a large chunk of information, e.g., a narrative, is being configured.

In contrast to focal information, elaborations are grounded in prior talk. As illustrated in excerpt (2), the speaker (Hb) uses yaʕni to frame her clarification of the concept fiʕira ‘cohabitation’, which she deems as culturally-specific:

(2)  Hb:  

1. fiʕira.  
   cohabitation  
2. yaʕni men șaknîn fi bēthum.  
   yaʕni two live in their house
Cohabitation, that is two [people] live in their house, have children, eat, drink, cohabitation.

In the examined corpus of written texts, 

yašni is far less frequent than in speech. Moreover, it is only used to frame elaborations. For example, in excerpt (3), the writer uses yašni to introduce a concretization of the more abstract idea he presents before:

(3) tab il-ṣamal is-ṣālih win-niyya is-salīma ?ahamm walla t-tašabbud? good the doing the good and the intention the good the most important or the worship? yašni wāhid xayyir wi-tayyib […] bass ma-bi-yiṣallīsh ha-yīthāsib ?iţ-zāyy? yašni someone good and kind […] but he does not pray he will be judged how?

All right then, are the good deeds and good intention the most important or the [religious] worship? That is, someone good and kind […] but who does not pray – how will he be judged [by God]?

It is easy to see how writing, as a pre-planned mode of discourse production, obviates the use of markers of online processing, such as the focus-framing yašni. A more intriguing question is why elaborations framed by yašni, which are clearly congenial to spoken interaction – specifically, to its incremental unfolding – occur in writing. A possible explanation may be that, although the situational motivation is lost, the structural and rhetorical functions of yašni are still retained in writing. That is, writers employ yašni as a general cohesion marker to expand previously introduced ideas (concepts, arguments, positions) in short fragments and in an often illustrative and ‘pedagogical’ style. In so doing, they also impart the text a flavor of casualness and naturalness, thereby underscoring the ideological motivation behind writing in the spoken idiom.


Session 3: Language Learning and Structure (Chair: Dorit Ravid)
Word order biases in adults and children: A silent gesture experiment
Maša Vujović, Gabriella Vigliocco & Elizabeth Wonnacott (University College London)

The distribution of word orders across languages is strikingly uneven, with SOV (Subject-Object-Verb) and SVO (Subject-Verb-Object) making up almost 90% of all languages with a dominant word order. Recent work suggests that SOV might be the default word order in language. Goldin-Meadow et al. (2008) found that adult participants preferred to use SOV in a silent gesture task, where participants describe simple events using gesture instead of speech. This preference was not only observed for speakers of an SOV language (Turkish), but also for speakers of SVO languages (English, Spanish, and Mandarin). In addition, subsequent work showed that the SOV bias disappears in favour of SVO when participants gesture about events with animate patients (Gibson et al., 2013, Hall et al., 2013). The effect of animacy is compatible with typological observations regarding word-order drift from SOV to SVO (Givon, 1979), and suggests a link between cognitive biases and linguistic structure. However, considering that adult and child language learning might be different (Hudson Kam & Newport, 2005), it is necessary to determine whether the same effects can be found in child learners.

We present two silent gesture experiments with children (6-year-olds) and adult controls. In Experiment 1, we used the same procedure as described in Goldin-Meadow et al. (2008). The events were presented as video animations, and involved either an animate patient (i.e., reversible events) or an inanimate patient (i.e., non-reversible events). Consistent with previous work, we found that adults were more likely to use SOV than SVO for non-reversible events (mean SOV: 55%) compared to reversible events, where the opposite was true (mean SVO: 83%). Children, however, largely produced a single gesture depicting the action in the video (mean 76%). It is possible that children produced uninformative responses due to a limited ability to incorporate into their productions the knowledge and informational needs of others. We therefore introduced a novel procedure in Experiment 2, which differed from Experiment 1 in that participants described the events to the experimenter, who had to guess the correct event from a choice of four. The task was designed such that providing the action gesture only was not sufficient for the experimenter to guess correctly, thus introducing a pressure on children to give responses that are more informative. Data collection for Experiment 2 is ongoing, but preliminary results show that Experiment 2 yielded fewer action-only responses from children, who appear to be using SOV and SVO equally (both for non-reversible and reversible events), and more than any other word order. Our results suggest that, given the right context, child learners, like adults, show a bias towards SOV compared with other unfamiliar word orders. Future work should aim to identify the specific mechanisms that link this early developmental bias with the distribution of word orders across languages.

References


SES Differences in the Structure of Child-directed Speech
Shira Tal & Inbal Arnon (The Hebrew University of Jerusalem)

One of the key findings in the literature on language acquisition is that socio-economic status (SES) impacts the input children receive: children from higher SES generally receive more input and higher-quality input than children from lower SES, a pattern that has cascading effects on language development [1-3]. To date, most of the work on SES-related differences in children’s input has focused on global measures such as the amount of utterances and words or lexical diversity [1-3]. Here, we ask if SES also impacts the structural organization of children’s input. Child-directed speech (CDS) is characterized by certain structural properties that facilitate language learning (e.g., [4]). These features, which impact the quality of the input, may also vary as a function of SES. One important structural feature of CDS is the frequent use of successive sequences with partial self-repetitions (see Example 1). These sequences, called variation sets (VSs), have been related to better learning outcomes in both naturalistic and experimental settings. Their proportion in CDS is correlated with the acquisition of verbs and multiword constituents [5,6] and their use in an artificial language led to enhanced word segmentation in adults [7], and better word learning in two-year-olds [8]. Here, we address two open questions about the role of variation sets: (1) Are VSs a unique feature of CDS or simply a characteristic of conversation? No study to date has actually compared their use in CDS and adult-directed speech, (2) is the use of VSs affected by socio-economic status: If SES impacts the quality of children’s input, as has been found for other linguistic measures, then we should see reduced use of variation sets in lower SES.

We addressed these questions in two corpus studies. Following [9], we defined variation sets as two consecutive utterances that share at least one word, excluding fillers, pronouns, wh-questions, and a set of function words. The first study compared the rate of VSs in child-directed speech (Gleason corpus, ages three-to-four, [10]) and adult-to-adult conversation (Santa Barbara corpus, [11]), a comparison not tested previously. We found that children indeed hear more VSs than adults do: both the proportion of words (POW) and the proportion of utterances (POU) were higher in CDS [POW: 33% vs. 17%, (t= -7.94, p<0.001), POU: 25% vs. 13%, (t= -6.85, p<0.001)]. The second study compared the proportion of VS in higher and lower SES mothers using the Howe corpus [12]. As expected, both POW and POU were higher in the higher SES group [POW: 39% vs. 30% (β = 0.08, SE = 0.03, p=0.01), POU: 33% vs. 26% (β = 0.06, SE = 0.03, p=0.03)]. These findings document the unique role of variation sets in child-directed speech; show that their use varies as a function of SES; and highlight the need to examine the effect of structural features of CDS, and not only global ones, on the trajectory of language learning.
Example 1
The following sequence, taken from the Howe corpus, is addressed to a two-year-old:

-Yes yes, he's got toes.
-Four toes.
-Have you got toes, Richard?
-Where are your toes?
-Show me your toes.
-Come and show me your toes.
-Where are your toes?

References
3. Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months, 16(2).
Differences between children and adults in the emergence of linguistic structure
Limor Raviv (Max Planck Institute for Psycholinguistics) & Inbal Arnon (The Hebrew University)
Limor.raviv@mpi.nl

Experimental work in the field of language evolution suggests that cultural transmission can lead to the emergence of linguistic structure as speakers’ weak individual biases become amplified through iterated learning [1]. These studies use diffusion chains (where the output of each learner is used as the input to the next learner) to show that unstructured artificial languages become more learnable and more structured over time as they are learned across generations. Similar findings have been reported in non-linguistic domains (see [2] for review). Interestingly, there is little evidence for similar patterns in children, despite such evidence being crucial for evaluating the role of learning biases in the emergence of linguistic structure. Children are the most frequent learners in the actual process of linguistic cultural transmission, and have been claimed to play a major role in the formation of linguistic structure (e.g., developing sign languages [3], creoles [4] and home-signing [5]), making the prediction that they will show stronger tendencies for structure emergence. However, children may differ from adults in their language learning skills, explicit prior knowledge and general cognitive biases. Data from young learners is crucial for understanding their role in language evolution and language change and the possible differences and similarities between child and adult language learning biases. Despite this, very few studies have look at iterated learning with children [6,7].

A recent study used a novel child-friendly paradigm to examine the emergence of linguistic structure with both children and adults, and found similar learnability curves despite adults’ better performance overall [6]. They also show that children can develop languages that are structured yet highly ambiguous, and similar to those created by adults under similar conditions [1]. While this study suggests that structure emerges also for child learners, it focused on languages that allowed homonyms and didn’t examine the crucial condition in which homonyms are not allowed and where compositional morpho-syntactic structure emerged in adults [1]. The current study extends prior work by testing this condition with both children and adults, using the same child-friendly task in [6] with an additional homonymity filter on the transmitted language (as in [1]) to create the expressivity pressure required for the emergence of compositionality [8].

We used mixed effects models to examine differences in trends of learnability and structure between the two age groups, using the same measures as in [1,6]. Results show that adults’ languages became more learnable over time, with errors significantly decreasing over generations (see Table 1). However, children did not show this pattern (see also Figure 1). Interestingly, there was no significant increase in linguistic structure over generations for either age group, although there were several instances of structured languages with partial compositionality in both child and adult chains. These results highlight possible differences in learning biases between children and adults, who outperform children overall (adults create more structured languages and make less mistakes from the very beginning).

However, these differences may be driven by children’s difficulty in learning the artificial language in general, especially after a relatively short exposure time (two exposures compared to six in [1]), a factor that may also explain why adults also did not show a significant increase in structure in this study. Therefore, in a new series of experiments we are currently testing this hypothesis by (a) increasing the exposure time, (b) reducing the number of semantic dimensions in the task in order to reduce the memory load for child learners, and (c) adding communication, a more nature way to create expressivity pressure [8]. While more work is needed to fully understand the cause for these differences, our results so far suggest that adults create more structure in this experimental paradigm, raising questions about children’ role in the formation of structure in newly formed languages [3,4,5], and highlighting the importance of testing child populations.
Table 1: Mixed effects model for transmission error

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.546729</td>
<td>0.027997</td>
<td>19.52794</td>
<td>&lt; .001 ***</td>
</tr>
<tr>
<td>Generation number</td>
<td>-0.03209</td>
<td>0.005813</td>
<td>-5.52026</td>
<td>&lt; .01 **</td>
</tr>
<tr>
<td>Age Group (Child vs. Adult)</td>
<td>0.121768</td>
<td>0.036671</td>
<td>3.320544</td>
<td>&lt; .01 **</td>
</tr>
<tr>
<td>Gender</td>
<td>0.030565</td>
<td>0.024057</td>
<td>1.270516</td>
<td>&gt; .1</td>
</tr>
<tr>
<td>Generation X Age group</td>
<td>0.023759</td>
<td>0.008113</td>
<td>2.928354</td>
<td>&lt; .01 **</td>
</tr>
</tbody>
</table>

Table 2: Mixed effects model for structure score

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.927903</td>
<td>0.177297</td>
<td>5.23361</td>
<td>&lt; .001 ***</td>
</tr>
<tr>
<td>Generation number</td>
<td>-0.0534</td>
<td>0.051505</td>
<td>-1.03687</td>
<td>&gt; .1</td>
</tr>
<tr>
<td>Age Group (Child vs. Adult)</td>
<td>-0.54456</td>
<td>0.214984</td>
<td>-2.53304</td>
<td>&lt; .05 **</td>
</tr>
<tr>
<td>Gender</td>
<td>-0.35224</td>
<td>0.207441</td>
<td>-1.69801</td>
<td>&gt; .1</td>
</tr>
<tr>
<td>Generation X Age group</td>
<td>0.109535</td>
<td>0.071931</td>
<td>1.522773</td>
<td>&gt; .1</td>
</tr>
</tbody>
</table>


Subjects and prepositions in acquisition:
The emergence of Preferred Argument Structure

The morpho-syntactic realization of arguments in a clause is a discursively governed syntactic phenomenon (Givón, 1983) which has been articulated, for instance, in the frames of Accessibility Theory (e.g., Ariel, 2001), and Preferred Argument Structure (e.g., Du Bois et al., 2003). It has been shown, for example, that the choice between zero, pronominal, and lexical nouns varies according to the referent’s level of accessibility (Ariel, 2001) and its role in the argument structure (Du Bois, 2003). The development of these principles in terms of language acquisition, however, has not been given systematic attention thus far. The present paper approaches the development of referential expressions in a syntactic and discursive perspective, accounting for their forms and functions, showing the emergence of the syntactic-discursive principles of Preferred Argument Structure and Accessibility Theory. Specifically, we investigate the realization of grammatical subjects and prepositional arguments in the peer talk of children (ages 2;0–8;0).

Grammatical subjects and prepositional arguments constitute the link between sentential argument structure and discursive needs in terms of referential expressions, and of construing the circumstances of the reported state of affairs. These linguistic elements pose a challenge to the young child as she learns the form-meaning correspondence between morphological variation, syntactic role, and discursive usage: The subject position is a core grammatical function (Andrews, 2007) that can be fully captured only when elements from different levels of grammar mutually determine each other (Lambrecht, 1994). Prepositions (e.g., into, for) constitute a comprehensive grammatical category expressing spatial, temporal and grammatical relations. They tend to be polysemic and lexically specified (e.g., work at vs. work for), and they tend to have semi-opaque paradigms in Hebrew (e.g., free im ‘with’, bound it-) and/or affix allomorphy (al-ay ‘on-me’, it-i ‘with-me’). That is, both categories require young children to pay attention to sentential (syntactic/semantic) and chain-of-reference (discursive) information. Research in child language shows that discourse-pragmatic information is integral to children’s use of language (Clark & Amaral, 2010; Salomo et al., 2010; Serratrice, 2008), with children’s argument realization reflecting such information as young as two years old (Allen, 2008; Huang, 2011; Orvig et al., 2010). Most developmental research accounts for each sentential ingredient on its own, describing and explicating its development. However, subjects and prepositional arguments both play a role in the argument structure of the clause. Thus, to better understand syntactic development, we propose to zoom out in order to examine the interplay between these grammatical categories within the clause as discourse dependent (Du Bois, 1987).

Preferred Argument Structure (e.g., Du Bois, 2003), and Constructional Preferred Argument Structure (Ariel et al., 2015), present two soft constraints on information structure, and specifically on the distribution of information in the clause: Quantity and Role. Quantity limits the amount of new information in argument positions of the clause core to no more than one lexical argument (that is, New or low-accessible argument). Role specifies where this lexical, New, low-accessible argument may appear in the clause, in terms of the syntactic relations of A, O, and S (transitive subject, transitive object, and intransitive subject, respectively); it is an argument-structure specific constraint which applies relative to the construction’s function. That is, Preferred Argument Structure provides a generalized theory of the link between syntactic relations and their discursive functions. Looking at how children express the arguments of a clause in a conversational context may shed a light on the development of the link between syntax and discourse. The present paper thus analyzes a peer-talk corpus of 56 children aged 2;0–2;6, 2;6–3;0, 3;0–4;0, 4;0–5;0, 7;0–8;0 respectively, engaged in triadic conversations (three triads per age group). We show that the emergence of grammatical subjects and prepositional arguments, and the diversification in their realization, reflect the development of the discourse-sensitive syntactic principles of Preferred Argument Structure.
References


Session 4: Language Processing (Chair: Bracha Nir)
In a recent paper outlining Cognitive Linguistics’ seven deadly sins, Dąbrowska (2016) draws attention to problems that must be addressed for the usage-based approach to continue to flourish. In this paper, we tackle two of the issues she addresses, viz. to take the Cognitive Commitment seriously and to study individual differences.

Dąbrowska (2016:483) argues that “it is time we stopped simply asserting that language is part and parcel of ‘general cognition’ (...) and started thinking more about how the analyses we propose fit in with what is known about [mental representations and processing]”. In addition, she calls on researchers to investigate individual differences. While individual variation naturally follows from a usage-based approach, in practice “most linguists, including cognitive linguists, do not look for individual differences, and tend to sweep them under the carpet when they find them” (485). We try to comply with her requests by examining the relationship between differences in linguistic experiences and anticipatory language processing.

Prediction-based processing is a fundamental cognitive mechanism, so much so that it has been stated that brains are essentially prediction machines (Clark 2013). This mechanism has been observed in language processing too: speakers generate expectations about upcoming linguistic elements and this affects the effort it takes to process them (for a recent overview see Kuperberg & Jaeger 2016). To a large extent, these expectations follow from knowledge about the patterns of co-occurrence of words, based on prior experiences. Given that people differ from each other in their experiences, expectations will vary across language users. This fact is rarely taken into account. Experimental measures of processing effort are related to cloze probabilities based on an amalgamation of other people's data, without addressing the question how representative these data really are.

We conducted two experiments with the same 122 Dutch participants. The participants belonged to one of three groups: Recruiters, Job-seekers, and people not (yet) looking for a job. These groups differ in experience with word strings that typically occur in the domain of job hunting (e.g. goede contactuele eigenschappen ‘good communication skills’). The groups do not differ systematically in experience with word strings that are characteristic of news reports (e.g. de Tweede Kamer ‘the House of Representatives’). For each register, we selected 35 word strings and we used these as stimuli in two experiments that yield insight into participants’ linguistic representations and processing: a Completion task and a Voice Onset Time task (VOT).

In the Completion task, participants were shown the first part of a string. They read aloud this cue and completed it by listing all things that came to mind. Mixed-effects models revealed that, on the news report items, the groups did not differ significantly from each other in how similar participants’ responses are to the complements observed in the Twente News Corpus. On the job ad stimuli, by contrast, all groups differed significantly (see Figure 1). The Recruiters’ responses were more similar to the complements observed in the Job ad corpus than the Job-seekers’ responses. The Job-seekers’ scores, in turn, were significantly higher than the scores of the Inexperienced participants.

In the VOT experiment, participants saw the cues once again, together with a complement selected by us. They were asked to read aloud this target word as quickly as possible. A measure of a participant’s own expectations proved to be a significant predictor of processing speed over and above word predictability measures based on amalgamated data (i.e. corpus-based surprisal estimates and group-based cloze probabilities). When participants had mentioned the target word themselves in the Completion task, they responded significantly faster than when they had not mentioned it (see Figure 2). Corpus-based lemma frequency proved to have a significant effect only when the target had not been mentioned by a given participant.

On the basis of these findings, we argue that it is worthwhile to go beyond amalgamated data whenever prior experiences form a predictor in models of language processing and representation. Our results show that there are meaningful differences to be detected between groups of speakers, and that a small collection of data
elicited from the participants themselves can be more informative than general corpus data. If cognitive linguists take their usage-based principles seriously, they ought to pay more attention to variation.

References


Figures

Figure 1: Mean score on the two types of stimuli for each individual participant, quantifying how similar each participant’s responses are to the complements observed in the corpora.

Figure 2: Scatterplot of the log-transformed corpus frequency of the target word (lemma), residualized against word length, and the Voice Onset Times, split up according to whether or not the target word had been mentioned by a participant in the preceding Completion task. Each circle represents one observation; the lines represent linear regression lines with a 95% confidence interval around it.
Abstract for the 3rd Conference on Usage-Based Linguistics, HUJ 2017

Processing typology and cross-linguistic variation in NP structure: An assessment of Hawkins’ efficiency predictions

Over the last three decades, John Hawkins has developed what is perhaps the most comprehensive and coherent framework for relating typological data to processing considerations of various kinds. Hawkins’ general proposal is that grammars tend to conventionalize morphosyntactic patterns in proportion to their degree of processing efficiency. This is reflected, for instance, in a cross-linguistic preference for word-order patterns that minimize domains of processing crucial dependency relations, and in limiting overt grammatical marking to relatively more demanding (i.e. unpredictable or structurally more complex) environments. Importantly, Hawkins (1986, 2004, 2014) has also repeatedly made predictions for the historical development of grammatical markers in different language types, such as VO and OV languages, in accordance with his efficiency principles.

The present paper is devoted to examining the typological robustness of one such prediction, relating to the internal structure of NPs. Across the world’s languages, noun phrases often contain elements in addition to the head noun (N) that can function as processing cues to the recognition (or on-line ‘construction’) of an NP, such as determiners, classifiers and related morphemes. Hawkins (2014) claims that such elements are more efficient in VO than in OV languages: As illustrated schematically below, an additional NP constructor C in a VO language can shorten the domain for the construction of the VP (V+NP), especially if N is delayed by intervening material (e.g. in AP-N sequences like the very delicious meal); in an OV language, by contrast, additional NP constructors lengthen this dependency domain, no matter where they occur in the NP:

\[
\begin{align*}
\text{V-NP processing domains in VO languages} & : & \text{V-NP processing domains in OV languages} \\
[V_{NP} N \ldots] & : & [N_{NP} C_{NP}] V_{VP} \\
[V_{NP} C_{NP} N \ldots] & : & [C_{NP} N_{NP}] V_{VP}
\end{align*}
\]

Hawkins thus predicts (among other things) that the grammaticalization of definite articles is a more productive process in VO than in OV languages, so that the synchronic distribution of articles is significantly skewed between the two language types.

While Hawkins briefly refers to Dryer’s (2005) WALS data in support of this claim, no rigorous evaluation according to current typological standards is offered. In addressing this shortcoming, we can turn to Bickel’s (2013) Family Bias method, which allows us to estimate whether the hypothesized processing advantage biases the development of genealogical units (families, stocks) in the predicted direction, and independently of macro-areal diffusion patterns. When used in conjunction with logit mixed-effects models (e.g. Jaeger et al. 2011), a robust and reliable assessment can be given: I will demonstrate that the results of both methods confirm Hawkins’ prediction at a global level, but that it becomes less convincing when more specific hypotheses are derived from it. For example, it does not seem to be the case that VO languages with ADJ-N order are particularly prone to develop articles, casting some doubt on the underlying assumption of strong domain-minimization pressures.

On a more general plane, I will also briefly comment on the extent to which Hawkins’ approach is compatible with core assumptions of usage-based linguistic theory.
References


How Does the Body Construct Linguistic Complexity?

Language is a communicative code that enables its users to express complex relations among entities, events, and states. Languages have developed sophisticated and complex grammatical devices for signalling these relations. How do these mechanisms originate and develop in a language? We cannot fully answer this question from spoken languages, since they are all thousands of years old or descended from old languages. However, sign languages of deaf communities can arise at any time and provide empirical data for documenting language emergence [1,2]. This paper presents evidence from a young sign language for the gradual emergence of linguistic complexity -- from basic to complex language forms -- through the gradually increasing complexity of body signals.

Sign languages make extensive use of manual, facial, head and body movements to encode a wide range of grammatical distinctions, e.g., [3,4,5,6]. Many of the signals generated by the body are functionally equivalent to prosodic features [7], [8], and provide an incisive tool for analysis of diachronic change in a sign language [9]. Our study traces the emergence of linguistic complexity through prosody in three generations of Israeli Sign Language (ISL) - a young language that originated only about 90 years ago [10].

Linguists working on contact languages, for example, suggest that certain aspects of language structure unfold in stages from basic to more complex over time [11,12,13,14]. In a young sign language, we are able to document the emergence of several such facets directly. (1) The first facet is prosodic segmentation into prosodic phrases, corresponding to thought units [15,16]. These units combine to form (2) constructions in which simple units are either adjoined loosely or are organised into (3) complex constructions in a hierarchical manner. Moreover, these complex units can be (4) embedded inside other complex units [12,17]. What structuring is present at the outset, and what is the order of emergence?

We studied two minutes of spontaneous narratives of 15 signers of ISL divided into three age groups (young=Y, middle=M, older=O). We divided all utterances into prosodic constituents, and coded the actions of every articulator of the body using established coding techniques. We identified 777 examples of constructions of varying complexities (topic-comment, dependency, coordination, listing, contrast) based on reliable markers of the body, which have been verified in the literature on ISL [7,18]. Finally, we analysed the ratio of each of these movements by calculating their frequency per two minutes of narrative and comparing their averages across the age groups.

Our findings, summarized in Table 1, indicate that there are no age-related differences in the frequency of the two basic properties of language: segmentation and topic-comment structure. All groups had a similar number of segmentation markings, i.e., bodily signals of prosodic boundaries, such up and down head movements. Similarly, we found no differences in the frequency of signals marking the most basic discourse relation: topic-comment. Thus, the encoding of segmentation and basic relations are already present at a very early stage of language, pointing to the primary significance of delineating thought units.

We did, however, find significant age differences in the frequency of encoding complex relations – relations between propositions (Y= 64%, M=57%, O=9% in Table 1). Younger signers (p< .05) produced a significantly higher number of forward head movements and opposite head tilts to mark dependent and coordinate constructions respectively. In contrast, the lack of explicit cues in the older age group means that complex relations between clauses need to be retrieved from context. Furthermore, we find that young signers (Y=56%) exhibit a higher number of embeddings - they more often incorporate one relation within another (e.g., coordination within dependency) – compared to older signers (O=2%).

To conclude, we show evidence that fundamental aspects of linguistic complexity emerge in specific stages. Signals of segmentation and basic discourse constructions (e.g., topic-comment) come first, followed by the emergence of more complex grammatical relations (e.g., coordination and dependency). Embedding of complex constructions within other complex constructions is last. In these cases, the bodily encoding of each relation is incorporated into a complex visual display (see Figure 1) which reflects functional complexity in a direct and compositional way through intricately coordinated articulations of different parts of the body.
Table 1. Distribution of signals marking different language aspects in 3 age groups of signers

<table>
<thead>
<tr>
<th></th>
<th>Older</th>
<th>Middle-aged</th>
<th>Young</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmentation</td>
<td>32</td>
<td>35</td>
<td>33</td>
</tr>
<tr>
<td>Topic-comment</td>
<td>30</td>
<td>38</td>
<td>32</td>
</tr>
<tr>
<td>Dependency</td>
<td>9</td>
<td>27</td>
<td>64</td>
</tr>
<tr>
<td>Embedding</td>
<td>2</td>
<td>42</td>
<td>56</td>
</tr>
</tbody>
</table>

Figure 1: Complex structure in narrative of a young signer. “[[[As I would watch the cartoons] [and connect them to the signed translation]], [my language developed.]]” The arrows trace the movement of the head, marking coordinated segments (a) and (b) with opposite head turns. Forward head movement simultaneously superimposed on the two coordinated units unifies them as a structure that is dependent on the main clause (c), when the head moves back to a more neutral position.

Session 5: Typology and Language Contact (Chair: Eitan Grossman)
A quantified notion of salience and its effect on contact-induced change
Eileen Waegemaekers and Umberto Ansaldo (The University of Hong Kong)

In a multilingual society where widespread bilingualism is the norm, language development and processing can be affected in a number of ways. One possibility is the emergence of a mixed variety that combines features of two or more languages. Building on work by Croft (2000), Mufwene (2001), Ansaldo (2009), Aboh (2015), it is assumed that language change at the individual level results from contact between different idiolects, and that features of those idiolects that the speaker has available, will compete for selection. In this paper we investigate whether there are inherent properties of linguistic features that can help us predict the outcome of the selection process.

Research on contact-induced language change suggests that for a feature to be replicated in the mixed variety it needs to have salient semantic content in either L1 or L2 (Aboh and Ansaldo, 2007; Siemund and Kintana, 2008). Following this line of thought and assuming that the primary locus of language contact is the mind, we hypothesize that when a speaker has the option to select features from different systems, the feature with a higher semantic salience will more likely get selected. However, the notion of semantic salience is hard to quantify in a principled manner, especially when required for cross-linguistic comparisons. In our approach, we quantify semantic salience as the relative contribution of an individual morpheme to the overall meaning of the sentence, coined as semantic load. A recursive neural network trained on learning abstract bilingual sentence representations of (Mandarin) Chinese and English is employed (Le and Zuidema, 2014; Hermann and Blunsom, 2014) that incorporates principles of distributional and compositional semantics. That is, the model uses frequency-based co-occurrence statistics and information on the compositional structure of sentences to learn word and sentence representations. Models that employ these properties have shown success at capturing meanings of words and phrases (Erk, 2012) and can enable us to quantify the semantic load of individual morphemes cross-linguistically.

The results from the model suggest that the semantic load hypothesis can account for several patterns found in situations of cross-linguistic influence and language contact. For example, one of the distinctive features of the mixed variety Singapore English is the widespread use of sentence final particles (Wong, 2004). Table 1 below shows results from the English-Chinese bilingual model and the ten word categories that get assigned the highest average semantic load. Sentence final particles get assigned a relatively high load, and accordingly, make up strong competitors and are therefore more likely to emerge in individual contact varieties, such as Colloquial Singapore English.

In closing, we also draw attention to the high semantic load of personal pronouns in both languages. As has been argued by Errington (1985), personal pronouns are pragmatically salient because of their referential nature, their reference to persons rather than objects and their indexicality. What the results from this model seem to suggest is that personal pronouns are also semantically salient, which can account for their recurrent role in linguistic changes (see e.g., Woolard, 2008, for examples). Our findings suggest that the quantified notion of semantic load can inform studies on contact-induced language change both at the individual and societal level. Moreover, it shows that computational models are not only useful in the domain of human-computer interaction but can also be used to directly inform linguistic research.
Table 1 Ten word categories with the highest average semantic load for English and Mandarin Chinese

<table>
<thead>
<tr>
<th>English</th>
<th>Chinese</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
<td>Semantic Load</td>
<td>Category</td>
</tr>
<tr>
<td>Personal pronoun</td>
<td>0.13</td>
<td>Pronoun</td>
</tr>
<tr>
<td>Modal verb</td>
<td>0.09</td>
<td>Subordinate clause marker</td>
</tr>
<tr>
<td>Possessive pronoun</td>
<td>0.08</td>
<td>Interjection</td>
</tr>
<tr>
<td>Coordinating conjunction</td>
<td>0.07</td>
<td>Sentence final particle</td>
</tr>
<tr>
<td>Adverb</td>
<td>0.06</td>
<td>Verb copula</td>
</tr>
<tr>
<td>Verb</td>
<td>0.06</td>
<td>Ba in Ba-construction</td>
</tr>
<tr>
<td>Interjection</td>
<td>0.06</td>
<td>Verb</td>
</tr>
<tr>
<td>Foreign word</td>
<td>0.05</td>
<td>Adverb</td>
</tr>
<tr>
<td>Noun</td>
<td>0.05</td>
<td>Proper noun</td>
</tr>
<tr>
<td>Existential ‘there’</td>
<td>0.05</td>
<td>Bei in Bei-construction</td>
</tr>
</tbody>
</table>

References


Thematic inflections in Navajo and Ket
Lukas Denk (University of New Mexico, Albuquerque)

Navajo and Ket are polysynthetic languages which exhibit verbal constructions where inflection and derivation are not expressed via distinct morphemes. Agreement and tense/mood/aspect are distributed constructions within the verbal complex and, further, are restricted in their use by the verbal lexicon. Thus, they have characteristics of both inflection and derivation, as well as characteristics of lexical items.

Vajda (2005, 2010) has pointed out the structural similarity between Navajo and Ket, and argued in favor of positing a genetic relationship between Na-Dené and Yeniseian languages. In this paper I will show that while in both languages inflection involves derivational and thematic information, the inflectional categories which do so are different. In Ket, agreement is lexically based, varying according to the verb, while TAM is mostly productive and generalizable (examples 1 and 3). In Navajo however, agreement is sensitive to the semantics and discourse status of the referent and generalizable across verbal constructions, while TAM is lexically restricted according to verb type (examples 2 and 4).

Usage-based approaches follow Bybee (1985), in noting the gradient character of the categories of lexicon, inflection, and derivation. Inflectional elements are relevant to the grammar; the meaning of these elements is relatively general and allows them to be applicable to a high number of stems. Derivational elements, instead, are relevant to meaning and function of the stem, and thus applicable to fewer stems.

Continua approaches provide a tool for accounting for the high variety of form-function mapping patterns across languages. However, I will argue that seeing derivation and inflection as two poles of a unidimensional continuum makes it difficult to address the similarities and the differences between the morphology of Navajo and Ket. In order to do so, I will argue that the continuum between lexicon and grammar is not to be understood as a unidimensional cline, but more as two dimensions which interact with one another. In this view, inflectional categories can still be fully inflectional – like agreement — but they can also serve as a tool for valence-change or be restricted to certain stems. The status of morphemes in Navajo and Ket might be better understood by means of a model that integrates two principles: 1) the interaction of two functions - lexical and grammatical – and 2) the interaction of two processing and accessing techniques - rule vs rote processing (Bybee 1985:207).
Examples

1) Ket: Different agreement pattern is used for different lexemes. In a), the agent is marked in slot 8 (du-) and -1 (-in), whereas in b), it is marked in slot 8 (du-) and 1 (an-).

a) \(du^{3}eji^{2}-\alpha^{3}\theta^{3}^{k}o^{3}-\alpha^{3}^{i}l^{3}^{2}e^{3}^{0}\in^{2}\)

\[\text{AN}_{3}^{\theta}-\text{TH-3AN}_{\alpha}^{\theta}-\text{TH-D-PT-call-AN.p_{A}}^{\theta}\]

'They call them.' (Vajda 2004:92)

b) \(du^{3}o^{3}n\theta^{3}^{k}a^{3}a^{3}^{l}q^{0}\)

\[\text{AN}_{3}^{\theta}-3\text{AN}_{\alpha}^{\theta}-\text{TH-1AN.p_{A}}^{\theta}-\text{sell}\]

'They sell them (animated entity off.)' (Vajda 2004:57)

2) Navajo: agreement dependent on the specificity of the referent (Mithun 2003:262-264)

a) \(yi^{3}\theta-\alpha^{3}zh\theta^{3}gh\)

IMPF-3s\theta-beautiful

'It (a horse or goat) becomes gentle, tame, tractable.'

b) \(ha^{3}\theta-\alpha^{3}zh\theta^{3}nee\)

\[\text{GEN}_{\alpha}^{\theta}-\text{be.lucky.NEUTR.IMPF}\]

'Things (weather, conditions) become pleasant, peaceful.'

c) \(a^{3}\theta-\alpha^{3}sh\theta-\alpha^{3}zh\theta\)

\[\text{UNSP}_{\alpha}^{\theta}-1\text{s}_{\alpha}-\text{eat.IMPF}\]

'I’m eating' (something unspecified)

3) Ket: Expressing imperative and past doesn’t add information to the lexicon (the shape of slot 2 might be also -in, which is thematic; however, the combination of the morphemes is straightforward).

a) \(ku^{3}i^{3}si^{3}-\alpha^{3}_{\theta}^{3}i^{3}^{2}-i^{3}j^{3}\)

2sj-call-DUR-ITER

'you.sg call.'

b) \(ku^{3}i^{3}es^{3}-\alpha^{3}_{\theta}^{3}i^{3}^{2}-i^{3}j^{3}\)

call-DUR.PAST-PAST/IMP-ITER

‘you.sg called.’

c) \(es^{3}-\alpha^{3}i^{3}l(d)^{3}i^{3}^{2}-i^{3}j^{3}\)

\[\text{call-DUR-PAST/IMP-ITER}\]

‘call!’

4) Navajo: TAM markers si- and yi-, both meaning ‘perfective’ add information to the lexicon/are lexically conditioned.

a) \(na^{3}di-si-sh\theta-\alpha^{3}lts^{3}\theta^{3}j^{3}\)

around-TH-PERF-1s-VAL-grasp/pinch

'I scratched around with my nails.'

b) \(yi^{3}\theta-\alpha^{3}sh\theta-\alpha^{3}lts^{3}\theta^{3}j^{3}\)

PERF-1s-\alpha-Val-grasp/pinch

'I plucked something.' (Young and Morgan 1992:634)

References


2 Abbreviations: AN = animate; DUR = durative; GEN = generic; IMPF = imperfective; s=singular

3 For the sake of clarity, the examples do not appear in their actual phonological form; this will be mentioned in the presentation.
Variation of the conditional marker in Võro
Helen Plado (University of Tartu, Võro Institute)

Võro is a Baltic-Finnic language spoken in Southeast Estonia. The Võro-speaking area is bordered by Russian in the east, and Latvian to the south. Traditionally Võro is regarded as a dialect of Estonian. According to the last census (2011), there are 87 000 Võro speakers. However, Võro has been under the strong influence of Standard Estonian for decades and today there are no monolingual Võro speakers.

Võro has three conditional markers: -(s)siq, -s, and -nuq (1). -(s)siq is the prototypical conditional marker in Võro. The nuq-marker is derived from the active past participle marker.

(1) Kõnõlõ-siq/kõnõlõ-s/kõnõl-nuq timä-ga.
   talk-COND (s)he-COM
   ‘I/you/(s)he/we/you/they would talk to him/her’

So far, there has been no usage-based research on the factors influencing the choice of form. However, some functional and areal variation has been noted: -nuq is mainly connected to eastern areas of Võro (Pajusalu et al 2009, Iva 2007), where the (more frequent) use of the form has been seen as showing influence from Russian, which also uses a past tense form to mark conditional mood; according to Keem & Käsi (2002), the nuq-form is used to express suggestion, concession, or reproach. Additionally it is claimed that the short s-marker is new and used nowadays primarily because of the influence of Standard Estonian (Iva 2002). The current study tests the validity of these remarks in corpus data.

This study addresses the following research question: what factors most influence the choice of the conditional mood marker in spoken Võro in the 1930s-1960s?

In order to answer the question, I have collected data from the Corpus of Estonian Dialects and from language data gathered during dialect interviews in the first half of the 20th century and published in two books. I applied Conditional Inference Tree analyses in R. The preliminary results demonstrate that the most important factor is verb type (auxiliary verbs are overwhelmingly marked by the shortest form -s), followed by tense and area. Remarkably, the nuq-form is most frequent not in the region neighbouring Russian-speaking areas, but in the region next to Latvian-speaking areas. Latvian has a similar construction (Kalnača 2014); hence, there is reason to assume that the use of the nuq-form is influenced by Latvian.

References

Co-expression patterns of nominal predication domains in Indo-Iranian

The grammatical co-expression of different semantic domains by the same structural coding means has been at the heart of several typological studies involving semantic maps and other methods (e.g., Van der Auwera & Plungian 1998, Hartmann et al. 2014). In the domain of nominal predication, similar studies have been used to argue for or against a privileged relationship between predicative possession and predicate location (e.g., Heine 1997, Baron & Herslund 2001, DeLancey 2002, Payne 2009), or to present a typology of the co-expression of core nominal predication and predicate location (e.g., Stassen’s 2013 WALS entry). Many of these studies treat co-expression as a binary variable: two domains are expressed by the same structural coding means or are not. Thus, these studies do not take into account the effects of intra-linguistic variation in the expression of functional domains. Moreover, some of these studies (e.g., Stassen 2013) limit their scope to the copula alone, effectively ignoring meaningful coding means such as argument flagging or word order.

This paper takes a more nuanced approach to the analysis of co-expression patterns. First, it treats co-expression as a continuous, rather than a binary, variable. Second, it takes into account an ensemble of structural coding means, and includes (following Clark 1978, Stassen 1997, inter alia) the identity of the copula, nominal flagging by case markers and adpositions, relative word order, and verbal indexation. Drawing on a set of published corpora in twelve Indo-Iranian languages, this paper reports two studies. The first study is concerned with the co-expression patterns of predicative possession and predicate location, and the second study follows Stassen’s 2013 question and is concerned with the relationship between the core nominal predication (equative, predicate attribute, proper inclusion) and predicate location.

Even when considering entire configurations of coding means together, and not just the copula alone, one finds clear and frequent co-expression patterns in naturalistic texts. The clauses in (1a-b), from Middle Persian, express predicate attribute and predicate locative by the same configurations of structural coding means. In both examples the same copular verb is accompanied by an unflagged NP and a prepositional phrase headed by andar “in, inside”. In examples (2a-b), from Gorani, the same verbal copula is accompanied by two unflagged NPs. Thus, (2a-b) exhibit the same configurations of structural coding means, but example (2a) expresses predicate attribute and example (2b) expresses predicate locative.

The textual data used for this study was scanned for clauses expressing core nominal predication, predicative possession, and predicate location (following definitions used in Clark 1978, Payne 1997:111–115, inter alia). All together, 200 – 500 tokens expressing nominal predication were collected per language. These tokens were annotated for the function they express and the structural coding means mentioned above. Following this, we compiled for each nominal predication domain the set of configuration of coding means which is used to express it in our data. When two nominal predication domains are never co-expressed, these sets would be completely disjoint, but often we find that pairs of sets overlap to varying degrees.

In both studies reported here, we measure how close to proper set inclusion are pairs of sets of configurations of coding means expressing predicative possession and predicate location (in the first study) and expressing core nominal predication and predicate location (in the second study). We find that Indo-Iranian languages vary in the degree in which core nominal predication and predicate location are co-expressed. In some languages the two domains are completely disjoint, but in others there are varying degrees of overlap. The sets of configurations of coding means expressing the two domains never completely overlap. We also find that predicative possession and predicate location are almost never co-expressed, despite the clear localist origin of some (or even most) possessor markers.

Cross-linguistic differences in co-expression patterns point, in turn, to variation in the type of language change processes in different nominal predication domains and in the type of synchronically active semantic extensions across the family.
(1a) mardom-ːn andar gumːn bːd hːnd
    man-PL in doubt be.PST be.PRS-3PL
“the people were in doubt” (Middle Persian, AWN)

(1b) was ruwːn ud frawahrːn andar ːn rːd bːd hːnd
    many soul and fravaːs in DEM river be.PST be.PRS-3PL
“there were many souls and fravaːs in this river” (Middle Persian, DK6)

(2a) dita-kaːš šīt biya
    daughter-DEF=3SG insane be.PST.3SG
“his daughter became insane” (Gorani, Mahmoudveysi et al. 2012: 98)

(2b) usā āsā faransa biya
    master then France be.PST.3SG
“at that time, the master was in France” (Gorani, Mahmoudveysi et al. 2012: 108)

References
Session 6: Round Table on the Acquisition of verbal Hebrew Morphology
The development of Semitic verb morphology has long challenged the psycholinguistic literature, given the derivational, semi-productive nature of the root and binyan verb pattern system (Berman, 1993). Verbs are lexical entities, with typical verb-related meanings (Kibrik, 2012), but in Hebrew their morphological components make critical contributions to their forms and meanings (Berman, 1985). Developmental accounts of Hebrew verb learning thus have to be based on empirical data regarding the type and token corpus distributions of root and binyan structures and functions, binyan temporal stems, and agreement markers. They need to point to the sources of evidence available to young learners regarding these components, and to explain the emergence and consolidation of the verb categories we observe in adult language.

Several questions are addressed in this context. One is the question of breaking into the verb system, that is, extracting root and pattern regularities, given the prevalence of Qal verbs with defective (irregular) roots, mainly in the modal cluster (Ashkenazi, Ravid & Gillis, 2016), which render early input and child speech highly opaque. Another question concerns derivational verb families, based on a single root shared by verbs with different binyan patterns, which organize the mature Hebrew verb lexicon and enable new derivation. How is the gap bridged between the core verb lexicon, characterized by Qal dominance, many singleton verbs, and a small number of highly coherent two-binyan families (Ravid et al., 2016), on the one hand, and the richness of the adult verb lexicon, where numerous derivational verb families vary in size, binyan composition, structural transparency and semantic coherence? A third question relates to the semantic relationships among members of a derivational family, and their effect on the patterns and rate of occurrence of verbs pertaining to the same family across the language learning years.

Adopting a usage-based approach to language learning (Tomasello, 2003), the thematic session seeks to provide a new account of Hebrew verb learning in native speakers from early childhood to adulthood. It is based on analyses of new Hebrew corpora comprising 500,000 words, including parent-toddler dyadic interaction, preschool children’s peer talk, children’s storybooks, and the written text production of schoolchildren and adolescents compared with adults. In addition to the Introduction, three talks comprise this session – (i) the emergence of the core verb lexicon in input to children and in child speech; (ii) SES differences in the consolidation of the verb lexicon in childhood; and (iii) the literate verb lexicon in Later Language Development.
References


The emergence of the core verb lexicon in input to children and child speech
Orit Ashkenazi, Ronit Levie, Galit Ben-Zadok & Dorit Ravid (Tel Aviv University)

The aim of this talk is to examine the characteristics of the core verb lexicon in Hebrew and determine the nature of the information children rely on in learning it. The studies comprised in the talk were grounded in Usage-Based acquisition approaches (Tomasello, 2009), which attribute major importance to Child Directed Speech (CDS) and to Child Speech (CS) as the source of distributions of lexical and morpho-syntactic units available to the child (Behrens, 2006).

To this end, two corpora were analyzed. One was a 370,000-word corpus of dense, tri-weekly 45-60 minute recordings of two Hebrew-speaking parent-child dyads between the ages 1;8-2;2 in natural interaction (Ashkenazi, 2015). Over 55,000 parental verb tokens and close to 8,000 child verb tokens were analyzed, yielding 521 root types and 684 verb lemmas in CDS, 224 root types and 259 verb lemmas in CS. Verb and root categories of CDS and CS were highly correlated within and across both dyads.

A second, 50,000-word corpus of 140 Hebrew children's storybooks had over 11,000 verb tokens, yielding 744 root types and 1048 verb lemmas. Storybooks are part of the ambient language children are exposed to, providing them with an initial, child-oriented source of literate Hebrew.

Parents and children were found to use a higher type frequency of full (regular) roots and a higher token frequency of defective (irregular) roots. Most (3/4) roots in this corpus derived singleton verbs, i.e., without same-root verbs in the corpus, and with the Qal pattern always dominating. The rest (1/4) of the roots derived highly coherent two-binyan families, with one member always dominating in token frequency (e.g., xibek 'hugged' 30 tokens, hitxabek 'hugged each other' 4 tokens). Most derivational families resided within the Qal-Nif'al-Hifil-Huf'al binyan subsystem. Larger derivational families were virtually absent from this database.

Storybooks revealed similar distributions, however with higher type and token frequencies of full roots. While Qal tokens dominated again, Pi'el tokens, representing a second binyan sub-system (Pi'el-Pu'al-Hita'el), were also prevalent. Most (2/3) roots in storybooks derived singleton verbs, with the rest (1/3) deriving mostly semantically coherent two-binyan families. However, storybooks also contained several larger (3+) morphological families, representing both binyan subsystems, in contrast to the sparser, more restricted families in spoken CDS and CS.

Findings thus suggest that CDS and CS verb characteristics facilitate early verb acquisition, so that the cradle of Hebrew verbs "starts small" in several ways (Elman, 1993). Morphological families in CDS and CS are few, so that verbs are learned as lexical items by high repetition, while roots and patterns are initially learned mostly through the temporal pattern structure of a single binyan (basic Qal). Small, semantically transparent root-binyan families highlight root-based and major transitivity relations (e.g., inchoative – causative), with the highly frequent member of the family ushering in the less frequent one.

While reflecting the same core characteristics of the Hebrew verb lexicon, storybook input had several enriching features. The higher prevalence of full root types and lower prevalence of defective root tokens indicate a lexically specific, more diverse lexicon. Moreover, storybook input exposes children to a more balanced binyan pattern structure than in spoken input, by adding Pi'el verbs from the newer, more productive binyan sub-system to verbs related to core Qal. Larger and less semantically coherent morphological families highlight root-based relations in storybooks, paving the way to the literate verb lexicon.
References


SES differences in the consolidation of the verb lexicon in peer talk across childhood
Ronit Levie, Hadus Hochman, Shirley Eitan & Dorit Ravid (Tel Aviv University)

Low socio-economic background is known to affect language development, especially impeding lexical learning (Rowe, Raudenbush & Goldin-Meadow, 2012). SES background is related to the amount and diversity of talk children experience from early on (Ravid & Zimmerman, 2016). From infancy, the low SES lexical repertoire lags behind that of more advantaged peers, showing slower trajectories of vocabularies growth (Black, Peppé & Gibbon, 2008; Hoff, 2013). The current study aimed to determine the effect of SES background on the development of the Hebrew verb lexicon in preschool children.

To this end, verbs were examined in the peer-talk of 36 monolingual, typically developing Hebrew-speaking preschoolers, aged 4-5 and 5-6 years respectively. This method offers a unique window on development, as children in peer interaction do not receive the kind of elaborative adult feedback that facilitates linguistic communication in mother-child dyads. The basic unit of the study was a 30-minute recording of a triad of same-age children in spontaneous play. Triads were selected for the study as the smallest group allowing children diverse opportunities for talk that could be captured on tape. Each age group consisted of three such triads, that is, nine children per age group. Half of the participants were from mid-high SES background, and half from low SES, following the criteria in Chiu & McBride-Chang (2006). Recordings were transcribed, and all verb morphological components were coded, including verb roots, binyan patterns and derivational verb families.

Findings revealed consistent quantitative and qualitative differences between the two SES corpora. The high SES corpus had about twice as many word tokens (12,605 vs. 5,706), verb tokens (2,585 vs. 1,387), verb lemmas (286 vs. 172) and root types (241 vs. 148) than the low SES corpus. In both SES, new lemma, verbform and verb token counts increased with age, but more steeply so in the high SES group. While both SES groups had higher type frequency of full (regular) roots and higher token frequency of defective (irregular) roots, full roots were more prevalent in the high SES corpus in both types and tokens. For both SES groups, Qal was the most prevailing binyan, followed by Hif’il and Pi’el. Again, for both SES, over 80% of roots occurred in singleton verbs, that is, with no verb derivational family. Also, 1/5 of the roots in both SES derived mostly semantically transparent two-binyan families, expressing basic transitivity modulations such as basic-inchoative, basic-causative, and inchoative-causative. However, the high SES transcripts had structurally and semantically more variegated singleton- and 2-binyan verb lexicons, within and across the two binyan subsystems.

Few studies have looked at children’s unsupervised peer talk as a source of information on language development, and none to date on children from different SES backgrounds. Results thus offer an additional perspective on lexical disadvantage in low SES preschoolers, pointing to a sparser and less diversified verb lexicon than that of their high SES peers, which foreshadows growing language and literacy discrepancies in the coming years (Levie, Ben Zvi & Ravid, 2016).
References


The literate verb lexicon in written text production across Later Language Development
Efrat Raz, Liat Hershkovitz, Ronit Levie and Dorit Ravid (Tel Aviv University)

The verb lexicon continues to develop in native language users throughout the school years, with exponential changes taking place in the literate lexicons of adolescents and young adults, concurrent with the advent of enhanced cognitive and linguistic capabilities (Nippold, 2016). The final talk in this thematic session examines the growth of morpho-lexical diversity in the Hebrew verbs produced during the period of Later Language Development, ages nine-adulthood (Berman, 2016).

The study investigated a 35,000-word corpus of 460 written narrative and expository Hebrew texts written by 230 4th graders (aged 9-10), 7th graders (12-13), 11th graders (16-17), young adults (19-20), and adults (Berman & Verhoeven, 2002). Text writers were all typically developing, monolingual, native Hebrew speakers from mid-high SES – that is, literate, but not expert writers. The corpus contained 7,298 verb tokens and 861 verb lemmas, based on 599 different roots.

Verbs and roots, including many new members of these categories, increased in number and diversity across the school years towards adulthood. Full (regular) roots prevailed across all age groups and in both genres, increasing with age and schooling. Qal accounted for half of the tokens and 1/3 of the types in 9-10 year olds (4th graders), declining in adults. In this youngest group, 2/3 of the verbs were singletons (decreasing to about half in adults), and about ¼ of the verbs were in semantically transparent two-binyan families, mostly expressing basic transitivity modulations. Derivational verb families increased to 1/3 of all verb lemmas in adults, with more variegated transitivity modulations, including truly passive verbs (Ravid & Vered, 2016). Less semantically coherent three- and four-binyan families emerged with age, with more complex morpho-lexical relations between the two binyan sub-systems (Qal-Nif'al-Hif'il-huf'al, Pi'el-Pu'al-Hitpa'el). These trends were more pronounced in narratives.

Hebrew verbs thus continue to develop in number and diversity in written text production across the school years. The features of the core Hebrew lexicon – reliance on basic Qal and singleton verbs - diminished with age and schooling, with increase in other binyan patterns. Written discourse by literate Hebrew users, especially narratives, was generally characterized by lexically specific full roots, with increasingly larger, semantically complex morphological families in older writers. This analysis of written Hebrew texts in Later Language Development (the school years) revealed a richly connected, root-and-binyan-based verb lexicon, with clearly established morphological and semantic categories. Unlike the core verb lexicon, this corpus showed more characteristics resembling those of the meta-linguistic view of Hebrew verb morphology prevalent in the non-developmentl linguistic literature.

References


The negation operator is not a suppressor of the concept in its scope.
   In fact, quite the opposite.
Israel Becker (Tel-Aviv University)

In psycholinguistics, the effect of the negation operator (henceforth, negator) on the activation levels of the concept in its scope is a controversial topic. Some argue that the negator automatically reduces the initially high activation levels of the concept in its scope to base-line levels or below, thereby assigning to the negator the role of a suppressor (e.g., Kaup et al. 2006; MacDonald and Just 1989). Others argue that the initial activation levels of the concept are not automatically suppressed by its negator, but sensitive to discourse goals. As a result, the negated concept is often retained in memory (e.g., Giora et al. 2005, 2007).

The current study resorts to natural speech in search of quantitative support for the Retention Hypothesis, rejecting the automatic Suppression Hypothesis.

The results of Giora et al.’s psycholinguistic experiments (2005, 2007) highlight an association between the mitigating function of a negator, modifying the negated concept, and its retention in memory: Giora et al. conducted both online (Giora et al. 2005, Experiment 1; Giora et al. 2007) and offline experiments (Giora et al. 2005, Experiment 3) using the same materials. In the online experiments Giora and colleagues showed that the initial activation levels of negated concepts are not any different from the activation levels of affirmative counterparts. In the offline experiments, comprehenders rated not good as less bad than bad and not bad as less good than good. Giora's results predict that, in natural speech, negative expressions will be construed by the speaker as conceptually and argumentatively weaker than an alternative in affirmative. Such cases are likely to manifest a highly accessible concept in the scope of the negator, to be marked by a zero anaphor $\emptyset$ (Ariel 1990).

To test Giora’s Retention Hypothesis, 400 instances were extracted from the spoken section of COCA (Davies 2008-). They all involved a discourse pattern, in which the speaker admits that NOT X (“not one that loves …”) is a weaker proposition than its affirmative counterpart (“I hate…”), by using an EMPHATIC CONNECTIVE (“in fact”), followed by THE OPPOSITE/CONTRARY. This discourse pattern is exemplified in (1), in which Larry King, the host of a nightly talk-show, and his guest, Donald Trump, discuss Trump’s upcoming divorce. Trump denies King’s insinuation that he would rather play the role of the ultimate playboy than have a monogamous relationship:

(1)  
[...] so I’m not one that loves the concept of divorce.  
   In fact, just the opposite $\emptyset$,  
   I hate the concept of divorce, I hate everything it represents.  
   There is nothing better than a good marriage. (27.7.1990)

Almost all instances of this discourse pattern contain a zero anaphor, $\emptyset$, in the scope of the OPPOSITE/CONTRARY. This zero anaphor, which refers to the concept in the scope of the preceding negator (i.e., the $X$ in the NOT $X$) attests to the highest possible accessibility of the prior negated concept (and, in fact, of any concept; see Ariel 1990), thus substantiating the retention (rather than the automatic suppression) of the concept in the scope of the preceding negator.
References


Competing motivations: changes in usage of constructions with saama + non-finite forms in written Estonian
Külli Habicht, Tiit Hennoste, Helle Metslang, Anni Jürine, Kirsi Laanesoo, David Ogren, Liina Pärismaa & Olle Sokk (University of Tartu)

The topic of our presentation is the variation of functions of the Estonian verb saama ‘get, become, can, may, will’ in constructions containing saama + different non-finite forms (e.g. ma saan minna ‘I can go’, ma saan minema ‘I will go’) in written Estonian from the 17th century until today. Thanks to its multifunctionality and high frequency of usage across different time periods, text types and registers, the verb saama is an effective vehicle by which to illustrate the mechanisms of change in language usage and the sociocultural factors influencing these changes.

Our research questions are (a) which functions are fulfilled by saama constructions in different text-types in different periods and (b) which sociolinguistic factors explain the changes in frequency of different saama-constructions. Our study belongs to the field of historical sociolinguistics and is based on usage-based perspective (Auer et al. 2015; Bybee 2010).

Verbs meaning ‘get, become’ are used in a variety of grammatical constructions in many European languages (e.g. Lenz, Rawoens 2012). Previous researchers have highlighted the usage of different saama constructions in Estonian (Penjam 2008; Tragel, Habicht 2012; Uiboaed 2013; Metslang 2016). We focus on the construction types that occur frequently in our data: 1) future construction with saama + supine (Ex. 1), 2) resultative passive construction with saama + passive participle (Ex. 2), 3) resultative impersonal construction with saama + passive participle (Ex. 3), 4) resultative construction with saama + active participle (Ex. 4), 5) modal construction with saama + infinitive (Ex. 5). The first two of these have taken root in Estonian on the basis of German werden constructions, the third and fourth are typical of both German and Estonian, and the fifth is purely Estonian.

Our data come from the language corpora of the University of Tartu. We took material representing different time periods, registers, and text types: 17th-18th century religious texts, 18th-century didactic-moralizing texts, didactic and fiction texts from the first half of the 19th century, fiction and popular non-fiction texts from the end of the 19th century, modern-day print media and fiction, computer-mediated comments and instant messaging dialogues. We randomly chose 100 sentences featuring saama-constructions from each text type (total 800 sentences).

Results. Our analysis reveals that the usage of saama constructions varies primarily by time period, not by text type or register.

Firstly, the general trend is one of movement from German-like usage toward typically Estonian usage. There is a clear difference between older (17th-19th century) and modern-day texts. In older texts, the future construction and the various resultative constructions are widespread, combining to account for 63–88% of all saama + non-finite form constructions in different text types, while the modal construction occurs relatively infrequently (3–25%). In modern texts, the modal construction dominates (80–88% in different text types), while the future and resultative constructions are much less common (totaling 7–11%). This coincides with two significant social changes: a) in the second half of the 19th century, the written Estonian language came into the hands of Estonians rather than Germans, and b) in the 20th century, Estonians adopted a more negative attitude toward German-influenced constructions.

Secondly, the usage of different German-influenced constructions (e.g. the saama future construction) varies across time. We explain this variation as the result of competition between language planning and the communicative needs of the linguistic community. Language planners have long endeavored to rid the language of constructions regarded as foreign, but these constructions have not completely disappeared; they may become marginal, but they remain available to be used when needed.
Examples

Ex. 1. Ne Haigkede pehle sahvat nemmat needt Kehdt pannema / sihs sahvat nemmat parrambax sahma. Auff die Krancken werden sie die Hände legen / so wirds besser mit ihnen werden. (REL)
‘They will put their hands on the sick, and then the sick will get better.’

Ex. 2. Kurjateggiatte käed said selja tahha seutud. (XIX-1)
‘The criminals’ hands got tied behind their backs.’

Ex. 3. Lähemal uurimisel − seda sai tehtud tänu prof. Jüri Kivimäe õhutusele − osutusid need leiud probleemseteks. (FICT)
‘Upon closer investigation − which got done thanks to the encouragement of professor Jüri Kivimäe − these findings proved to be problematic.’

Ex. 4. ...Mart, saad sa söönd, wöta hobbose ja äkke, ja tulle järrele. (XIX-1)
‘Mart, when you get done eating, take the horse and the harrow and come after me.’

Ex. 5. Neid asju saan stipendiumi tõttu nüüd lihtsalt paremini teha. (NEWS)
‘I can do those things better now because of the stipend.’

References


Corpora

Abbreviations: FICT – modern-day fiction; NEWS – modern-day printed media; XIX-1 – texts from the first half of the 19th century; REL – 17th-century religious texts.
The role of agentivity in the Spanish causative hacer construction

In the Spanish causative hacer construction ‘to make someone do something’ (henceforth, ChC), the causee can be expressed with the accusative clitic ‘lo/la/los/las’ or with the dative clitic ‘le/les’ as seen in the example in (1).

(1)  Lo /Le hice probar la comida.
    CL.ACC/DAT made.1SG to-try the food
    ‘I made him.ACC/DAT try the food.’

The question that arises is: what determines the accusative-dative case marking in the ChC? Previous studies claim that the transitivity of the embedded verb in the ChC determines case (Aissen and Perlmutter, 1976; Rosen, 1990, *inter alia*). That is the causee is assigned accusative or dative case depending on whether the embedded verb is intransitive or transitive respectively. However, transitivity does not sufficiently account for case marking in the ChC since there are sentences, as seen in (2), where the intransitive verb pensar ‘to think’, reir ‘to laugh’, sonreir ‘to smile’, and llorar ‘to cry’ are used with dative case marking.

(2)  El poeta agregó que Monsiváis le hizo pensar y Elena
    the poet added.3SG that Monsivais CL.DAT made.3SG to-think and Elena
    Poniatowska, su amiga, le hizo reir, sonreir y a ratos casi
    Poniatowska his friend CL.DAT made.3SG to-laugh to-smile and at whiles almost
    llorar.
    to-cri
t    ‘The poet added that Monsivais made him think and Elena Poniatowska, his friend, made him
    laugh, smile and at times almost cry.’

(Example from the Corpus de Referencia de Español Actual)

A second approach proposed is the Semantic Approach (Shibatani, 1975; Ackerman and Moore, 1999; *inter alia*) which posits that there are several semantic contrasts that determine accusative-dative case marking in the ChC. Following the Semantic Approach, I conducted a lexical semantic study of case marking in the ChC using data from two corpora: the Corpus de Referencia del Español Actual and the Corpus del Español. Corpus data comes with rich context that exposes many factors that come into play in language use and variation. Based on my findings, I argue that the semantic factor of agentivity, measured on a scale, influences case marking in the ChC. Agentivity is directly related to coercion, volitionality, and animacy; factors that have all been posited in previous research as semantic contrasts that determine case marking in the ChC (for coercion see Shibatani, 1975; Strozer, 1976; for volitionality see Ackerman & Moore, 1999; and for animacy see Finneman, 1982). I define agentivity as: a predicate entails agentivity if there is an action performed by the subject and/or the predicate entails volitionality of the subject. The diagnostics I used for agentivity are: the added conjunction test (*y X lo hace/hizo + modifier ‘and X does/did it + modifier’) and the volition test (adding adverbial phrases that express volition to the sentence). After analyzing the data for agentivity, the results suggest there is a correlation between higher degrees of agentivity and accusative case marking and lower degrees of agentivity and dative case marking. For example, in (2), the causee is forced to do the action, that is, there is higher agentivity on part of the causer and lower agentivity on part of the causee. Therefore, the causee appears with accusative case marking and the dative case marking is not acceptable.

(3)  Lo /*Le hice probar la comida a fuerza.
    CL.ACC/DAT made.1SG to-try the food at force
    ‘I forcibly made him.ACC/*DAT try the food.’

Similarly, in (2), there is lower agentivity on part of the causer and higher agentivity on part of the causee, and thus, the causee appears with dative case marking. My results also suggest that the dative, contrary to expectations, is the unmarked case for the ChC. The agentivity scale I present is not unconditional as there are several factors that contribute to case marking which require further study. Finally, I argue that my hypotheses support cross-linguistic and theoretical work on causation (Croft, 2002) as well as the theories on transitivity in the field of general linguistics given that agentivity is a component of transitivity (Hopper & Thompson, 1980).
Bibliography


Corpus de Referencia de Español Actual, REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [en línea]. *Corpus de referencia del español actual*. <http://www.rae.es>


Effects of Immigration on L1 Jamaican Creole and L2 Standard American English Verb Forms
Taryn Malcolm and Loraine K. Obler (Speech-Language-Hearing Sciences, CUNY Graduate Center)

Few studies have examined the specific morphosyntactic forms employed among Jamaican Creole (JC)/Standard American English (SAE) speakers living in the U.S. Yet we might expect that their bilingualism results in transfer from the first language (L1) to the second language (L2) (e.g., Montrul, Bhatt, & Bhatia, 2012) as well as L2 influence on L1 (e.g., Pavlenko and Jarvis, 2002). All individuals educated in Jamaica will have learned standard English verb forms, but immigration to the U.S. and immersion in its culture can be expected to influence their SAE further, perhaps at the expense of the JC.

The purpose of the current study was to determine how changes in tense markers (e.g., JC past tense “did”, future tense “a go”) manifest in adult bilingual speakers of JC and SAE living in the United States for no fewer than 5 years. Verb tense markers were our focus because, unlike SAE, JC lacks bound inflectional morphemes and instead uses pre-verbal markers to indicate tense (e.g. JC: (H)im did cook rice an peas = SAE: He cooked rice and peas; JC: Mi a go shop = SAE: I will go shopping.)

This study explores the role of language use, motivation, and attitudes towards the two languages in retention or attrition of the L1, as well as how these factors impact performance in L2. These factors are crucial in determining the proficiency level of each person’s L1 and L2 and how differences in the perception of each language interacts with usage patterns following immigration to the United States.

Six participants have been tested to date (average current age 43.1; average age at immigration 30.5) All had at least some secondary schooling and had not completed a college degree. Verb forms were elicited by asking participants to orally repeat 30 sentences in each language with past tense verbs and 30 with future tense verbs. The stimuli were presented via digital recordings of native speakers. Participants’ responses were recorded and transcribed, then repetition-correctness of marked JC and SAE verb features was scored.

Repetition of verb forms was quite good for JC sentences, except by one participant who correctly repeated the ‘a go’ form in only 54.3% of instances (the rest of the participants repeated the forms as presented on average 95.4% of the time.) Repetition was poorest for the SAE past tense, with participants ranging from 31.1% correctly repeated to 93.1%; average: 68.6%). Greater daily use of English correlated with better repetition of SAE verb markers and worse repetition of JC forms. Attitude toward each language at work was more important than attitude toward language use at home in predicting repetition success.

Morphophonological constraints on syllable-final consonant clusters clearly enter into the poor repetition of past-tense SAE inflections overall, however that they should interact as well with reported percent of time using each language and attitudes toward language at work is of interest. Maintaining L1 verb forms appears to be linked to attitudes towards the respective languages, as does, mutatis mutandis, employing L2 ones.

References
Daily Use and Repetition Errors

Attitude Toward Language at Work on Repetition Errors

“I find speaking Patois valuable at work.”

“I find speaking English valuable at work.”
A complex clausal structure and a discourse marker – the case of Hebrew ‘azov (lit. ‘let go’)
Hilla Polak-Yitzhaki (Haifa University and Gordon College of Education)

Hebrew utterances such as

...‘azov shedoktorat, let go.IMP.2.SG.M that-PhD ...
...let go of the fact that a PhD, ze mashehu mexusbad, is something respectable,

are considered bi-clausal constructions consisting of a main clause ‘azov (imperative ‘let go’) consisting of a predicate in the second person imperative form, and a subordinate complement clause opening with the complementizer she- ‘that’. However, when such tokens are investigated within an interactional linguistics approach (Selting and Couper-Kuhlen 2001), ‘azov – the part traditionally considered the “main clause” – is shown to function as a ‘projector construction’ (Auer 2005, Hopper and Thompson 2008, Günthner 2011). A synchronic analysis of all tokens of the verb ‘azzav (‘leave, let go’) – 89 tokens – throughout an audiotaped corpus of 243 informal Hebrew interactions (over 11 hours of talk) suggests that this verb has evolved into a discourse marker.

More than half the tokens (54%) are in the imperative (41 tokens) or future (7 tokens). 50 tokens (56%) are employed nonliterally and occur in the following constructions:

1. ‘azav+lexical NP (4 tokens) instructs the hearer that although there are different aspects to the topic discussed, the speaker chooses to focus only on one aspect.

2. ‘azav+Pro (22 tokens)
   a. ‘azav+1st/3rd P ACC Pro (16 tokens) – a metaphorical use related to the extralingual world: instructs the hearer to let go of a discussed referent, or reports (not) letting go of it.
   b. ‘azav+2nd P ACC Pro (6 tokens) – a prototypical interpersonal discourse marker (Maschler 2009): rejects the interlocutor’s idea while expressing the speaker’s stance towards it

3. ‘azav+she- (‘that’)/question word (5 tokens) – a turn-medial ‘projector construction’ intensifying a previous assertion by instructing the hearer that it is even stronger than a following assertion framed by the ‘azov she- construction:

1 Iddo: ..‘ani be’ezrat hashem, I in-help-of the-name ..I with the help of God,

2 ...
...xatsi shana, half year ...
for half a year,

3 ..lo kibalti ‘afilu tsav, not received.1.SG even order ..haven’t even received an order to go on reserve duty.

4 ...‘azvu shelo hayiti bemilu'im, 
...let go.IMP.2.PL that-not was.1.SG in-reserve military service ...
...let go of the fact that I haven’t been on reserve duty.
The fact that Iddo hadn’t been on reserve duty for six months is less surprising than the fact he hadn’t even received an order to go there, since it is possible to request deferral of such an order. In line 4 after presenting the more surprising information (lines 1-3), regarding not receiving the order, Iddo employs an utterance opening with ‘azvu she- conveying the less surprising information. In lines 5-6 Iddo repeats the information from lines 1-3, which, in light of line 4, becomes even more surprising.

4. ‘azav in its own separate intonation unit (19 tokens)
   a. ‘azov/l’azvi (‘let go’ [IMP.2.SG.MASC/FEM] /ta’azvi (‘let go’ [FUT.2.SG.FEM]) in a continuing intonation contour (10 tokens) in responsive position: projects disagreement, or signals a shift to a new discourse topic.
   b. ‘azov/l’azvi (‘let go’ [IMP.2.SG.MASC/FEM]) in a separate sentence-final intonation contour (9 tokens) – dismisses the hearer’s idea with no elaborations.

Elaborating on these findings, I will present a grammaticization (Hopper 1987) path: an evolvement from the literal use of concrete leaving, through a nonliteral use of abstract-mental leaving, to a metalingual use (Maschler 2009) of leaving a discourse topic (a textual DM), or expressing rejection (an interpersonal DM).

Thus, not only does this study undermine traditional analyses by revealing the inadequacy of concepts such as ‘bi-clausal’, ‘asyndetic clause’ and ‘parenthetical’, it also demonstrates how the different ‘azav constructions and their functions emerge from the interlocutors’ recurrent social actions in discourse.

References


Prepositions (e.g., into, for) constitute a comprehensive grammatical category expressing spatial, temporal and grammatical relations. They require young children to pay attention to the preposition’s meaning / function in both sentential (syntactic/semantic) and chain-of-reference (discursive) contexts. Moreover, prepositions tend to be polysemic, with several different functions for a single prepositional form (Tobin, 2008), and they are often specified for particular lexical contexts, resulting in semantic nuances (e.g., work at or work for). Hebrew prepositions add another level of complication for young learners by being inflected for gender, number, and person when complemented by a pronoun (e.g., al-ayix ‘on-you,Sg,Fm’). Inflected prepositions may undergo stem change (e.g., free im ‘with’, bound it-) and / or affix allomorphy (al-ay ‘on-me’, it-i ‘with-me’), resulting in semi-opaque paradigms.

Prepositions thus pose lexical, semantic, morpho-phonological, syntactic and discursive challenges for language acquisition. Nonetheless, as a frequent category in spoken language (Johnson & Slobin, 1979), some basic prepositions are acquired as early as age two years, side by side with the emergence of syntactic structure (Berman, 1993). The current study hypothesized that the emergence and diversification of prepositions in the course of acquisition are a function of the interplay between frequency and functionality, as realized by entropy (Martin et al. 2004). To confirm this hypothesis, all 3058 preposition tokens were identified in a corpus of peer talk by 60 typically developing, monolingual Hebrew-speaking children in four age groups (2;6–3;0, 3;0–3;6, 3;6–4;0, 4;0–5;0 respectively). They were engaged in 45-minute triadic conversations at their preschool (five triads per group). Each token was coded for its lexical semantics, morphological, and syntactic properties.

Findings showed that with age, preposition tokens and types grew more numerous, more morphologically complex, populating more diverse syntactic and discursive contexts. Their developmental path confirmed our hypothesis, showing that the category of Hebrew prepositions emerges and expands as a factor of both frequency and function, as realized in the entropy of each preposition and function. The entropy of a preposition is an estimate of the amount of information it carries: The greater the number of functions a preposition has, the greater its entropy. The entropy of a function is an estimate of the prepositions that serve it. The greater the number of prepositions serving a function, the greater the entropy of the function tends to be. The estimates of the entropy of functions and prepositions explain the expansion and diversification of prepositions age. On the one hand, a function is coupled with a small number of forms at the beginning of acquisition, and form varies with age. On the other hand, a preposition is coupled with a small number of functions at the beginning of acquisition, and function varies with age. This entropy-based explanation interacts with frequency-based results, which show that the rise in frequency is correlated with morphological complexity and with syntactic functionality. Having emerged, the category diversifies in older children (as shown by Morgenstern & Sekali, 2009), with new prepositions, and new functions of old prepositions, referring to more abstract relations.
References


Grammatical Error Correction (GEC) is a challenging research field, which interfaces with many other application areas of linguistics. The field is receiving considerable interest recently, notably through the GEC-HOO (Dale and Kilgarriff, 2011; Dale et al., 2012), CoNLL shared tasks (Kao et al., 2013; Ng et al., 2014). Within GEC, considerable effort has been placed on system evaluation, which is notoriously difficult, much due to the many valid corrections each source sentence may have (Tetreault and Chodorow, 2008; Madnani et al., 2011; Chodorow et al., 2012; Dahlmeier and Ng, 2012).

An important criterion in the evaluation of GEC systems (henceforth, correctors) is their ability to generate corrections that are faithful to meaning of the source. In fact, many would prefer a somewhat cumbersome or even an occasionally ungrammatical correction over a correction that alters the meaning of the source (Brockett et al., 2006). As a result, often when compiling gold standard corrections for the task, annotators are instructed to be conservative in their corrections, e.g. in NUCLE (Dahlmeier et al., 2013) and the Treebank of LL (Yannakoudakis et al., 2011). A recent attempt to formally capture this precision/recall asymmetry has been the shift from using $F_1$ to $F_{0.5}$, where Precision is emphasized over Recall, as a common evaluation measure in GEC (Dahlmeier and Ng, 2012).

However, favoring precision over recall may lead to reluctance of correctors to make any changes. Using a single reference correction (a common practice in GEC) compounds this problem, as systems are not only penalized more for making an incorrect change, but are often penalized for any correct change not found in the reference.

We present results that indicate that current state of the art systems suffers from over-conservatism. Evaluating the output of 15 state of the art GEC systems that participated in the recent CoNLL2014 shared task, we find that all of them substantially under-predict corrections relative to the gold standard.

We pursue two approaches to address this gap, proposing improved reference-based protocols to decrease penalties of valid correction. Thus, avoiding over-conservatism. First, we study the effect of increasing the amount of references (henceforth, $M$). While previous evaluation explored the case of $M = 2$, no empirical assessment has been carried out, of its sufficiency or its added value over $M = 1$. In our experiments we estimate the number of corrections necessary to cover the bulk of the distribution of possible corrections. We then consider two measures for assessing the validity of a proposed correction relative to a set of references, and characterize the distribution of their scores as a function of $M$. We find that assessment based on $M = 1$ or $M = 2$ dramatically under-estimate the true performance. We conclude with an analysis of the statistical significance of these evaluation protocols.

Second, we pursue an alternative way by developing a semantic measure based on the similarity of their semantic structures. We define a measure, using the typology inspired UCCA scheme (Abend and Rappoport, 2013) as a test case. In order to assess the feasibility of this approach, we annotate a section of the NUCLE parallel corpus. Our results support the feasibility of the proposed approach, by showing that semantic structural annotation can be consistently applied to LL and that manually com-
Piled references do indeed have similar semantic structures to those of their source sentences.

The two approaches address the insufficiency of using too few references from complementary angles. The first attempts to cover the bulk of the distribution of possible corrections for a sentence, while the second uses semantic instead of string similarity, in order to abstract away from some of the formal variation between different valid corrections.

References


Figure 1: Mean accuracy values for perfect correctors with different number of references. Lines shows accuracy based on exact match and exact index match, match off the same words were deleted added or changed.


Table 1: Different UCCA distance measures between LL and corrected parallel paragraphs. The scores are quite similar to the IAA, which means distances are indeed low.

<table>
<thead>
<tr>
<th>Annotator</th>
<th>Edit</th>
<th>Fully aligned</th>
<th>Top down</th>
<th>Token analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same</td>
<td>323.16</td>
<td>0.78</td>
<td>0.79</td>
<td>0.85</td>
</tr>
<tr>
<td>Different</td>
<td>209.55</td>
<td>0.85</td>
<td>0.87</td>
<td>0.89</td>
</tr>
</tbody>
</table>
Following Haspelmath et al. (2014), the present study aims to form a typology of coding patterns of causal-noncausal verb pairs in German speech islands – a group of minority heritage languages (a) spoken outside of and removed from their original geographic and cultural territory, (b) which are surrounded by majority substrate (or superstrate or adstrate) languages with which they experience prolonged and influential contact and (c) their speakers are at least bilingual, which affects their use of their heritage language. There is one major advantage to studying speech islands – they are, to quote Schirmunski (1930), a “linguistic laboratory”, due to the fact that they all stem from a single source, but have changed differentially over a relatively short period of time, which allows one to investigate the effects of language contact in real time.

According to Croft, “the primary framework for understanding event structure and verb meaning is causation” (1990:49), that is, verbs can be categorized according to the basic meaning of their core event, and whether or not there is an inherent sense of cause to it. Haspelmath et al. (2014) use the term causal-noncausal alternation and this terminology is adopted in this study as well. These alternations are coded by various morpho-syntactic means (i.e., synthetically or analytically), termed causative (that is, showing extra marking on the causal verb in the pair), anticausative (showing extra marking on the noncausal verb in the pair), equipollent (i.e. having completely different forms), or labile (identical form is used for both verbs in the pair without any overt marking).

It is often the case that typological studies are based on samples that explicitly aim to be balanced and as free of genetic, geographic and demographic biases as possible. In this study, a different approach is taken, one which directly targets the effects of both genealogy and language contact as influential factors in language change; the present study develops a micro-typology of German speech islands, a group of languages that are genetically related to one another, and are in continued contact with other, distinct languages. Preliminary research, in which 12-14 causal-noncausal verb pairs were examined in four languages (Standard German, Pfälzisch, Pennsylvania German and English), shows interesting changes in the coding types of causal-noncausal verb pairs in PG, with a general shift towards labile coding, possibly under the influence of English, a distinctively labile language (Table 1). Additionally, anticausative coding, while relatively rare in SG and Pf, is completely absent in PG and English, again hinting at the influence of contact (Table 2).

As stated, the goal of this study is to map out the coding types German speech islands employ in order to mark causal-noncausal alternations; furthermore, it is aimed to discover the role different factors play in determining them; usage motivated internal developments, universal tendencies, and language contact are all possible contributors to the observed changes in the data. Thus far, there is strong evidence leading to contact being an important factor, and this study is an on-going project hoping to support – or refute – this hypothesis.
### Table 1

<table>
<thead>
<tr>
<th>Standard German</th>
<th>Pfälzisch</th>
<th>Pennsylvania German</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>open (causal/noncausal)</td>
<td>öffnen/öffnen sich</td>
<td>uffmache/uffmache</td>
<td>open/open</td>
</tr>
</tbody>
</table>

**Anticausative** Equi-pollent | Labile | Labile |

### Table 2

| Anticausative | 2/13 | 2/12 | 0/13 | 0/14 |

### References


The ordering of main and temporal adverbial clause in Czech
Michal Láznička (Charles University, Prague)

The placement of adverbial subordinate clauses relative to a main clause is flexible in Czech. This pattern is quite common cross-linguistically (Diessel 2001). However, authors, working almost exclusively with introspection, do not agree on the main factors that influence clause order in Czech (e.g. Havránek & Jedlička 1970). The aim of the present paper was to determine the amount of variation in the placement of four types of temporal adverbial clauses in Czech and to identify factors that influence clause order. A sample of sentences containing one of the four connectives než, předtím než (both ‘before’), poté co, až (both ‘after’) was extracted from a corpus of written Czech. 500 sentence were sampled for each of the connectives. Non-temporal clauses were filtered out and a total of 1204 sentences was coded for structural variables (length, complexity, status of head clause, subject identity, tense and aspect of head and subordinate clause verb) and a single semantic conceptual variable, temporal iconicity. Temporal iconicity is characterized as a diagrammatic mapping of the order of referents and events in discourse to the order of events in reality. There is no need to explicitly indicate that the drinking preceded the eating in 1). However, it is possible to override this principle by using functional elements, such as the connective než ‘before’ in 2), which makes it possible to reverse the order of events in discourse relative to reality.

1) Vypil sklenici vody a snědl preclík. ‘He drank a glass of water and ate a pretzel.’
2) Než snědl preclík, vypil sklenici vody. ‘Before he ate a pretzel he drank a glass of water.’

Diessel (2008) has shown that iconicity remains a significant factor in the placement of English temporal adverbial clauses, even in the context of grammaticalized connectives that explicitly code temporal relations between events.

On the whole, adverbial clauses tend to follow the head clause. Furthermore, three factors were identified to influence the position of the adverbial clause in logistic regression. Iconicity, status of the head clause (main or subordinate clause), and relative length of the subordinate clause were identified as significant predictors of clause order in the analysis. Longer clauses as well as clauses the heads of which are themselves a dependent of another clause are more strongly associated with postposition. In addition, ‘before’ clauses are also more strongly associated with postposition. This may be interpreted as an effect of iconicity: other things being equal the ‘before’ clauses prefer postposition; just as the event expressed by these follows the main clause event in reality. To follow up on this finding, preliminary results of an exploratory study aimed to assess the role of iconicity in sentence processing will be presented. In a self-paced reading experiment participants will be presented sentences beginning with a ‘before’ or an ‘after’ clause. I will look for differences in reading times at clause boundaries. If iconicity plays a role in processing, ‘after’ clauses in initial position will facilitate processing, compared to (non-iconic) ‘before’ clauses.

References
Causal–noncausal verb alternations in Yiddish in comparison with German and Russian
Elena Luchina-Sadan (Hebrew University of Jerusalem)

1. Overview
This paper provides a corpus-based analysis of the basic alternations in verbal system. As Nichols et al. (2014) demonstrates particular languages tend to prefer a particular morphological strategy of coding the difference between causal and non-causal events. Besides causative (the transitive verb bears a transitivizing morpheme), anticausative (the non-transitive verb bears a detransitivizing morpheme), there are two more types: equipollent (both forms are marked) or labile (the same form is used as transitive and intransitive). Haspelmath et al. (2014) show that coding asymmetries corresponds with frequency, that is, the more frequent element in a given causal-noncausal pair (e.g., ‘raise’ vs. ‘rise’) is normally coded with a shorter form, while its partner is coded with a longer (or overt) form. Testing these ideas on Yiddish data gave interesting results.

This analysis is based on dictionary data and the Corpus of Modern Yiddish (ca. 4 million words), which includes several different genres, mainly classical literature, Soviet prose, and modern secular press.

2. Transitivizers and detransitivizers in Yiddish
There are two detransitivizers in Yiddish: the reflexive pronoun zikh and the originally passive construction PTCP + vern ‘become’ The former is general and widespread, while the latter is used for the verbs denoting extreme change of the state (Fedchenko 2016).

As for transitivizers, inseparable prefixes can form ba-ru-ik-n ‘calm down (tr.)’ ← ru ‘calmness’, far-leng-er-n ‘lengthen’ ← lang ‘long’. But with verbal stems not only ba- and far-, but probably the 4 other inseparable prefixes work as perfectivizers. Cf. the corresponding study of German prefixes (Zeller 2001).

3. Results of the corpus experiment
The questionnaire proposed by Haspelmath et al. contains 20 verbs forming a scale from more spontaneous to less spontaneous. In the table the order of the forms and labels corresponds to their frequency.

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Yiddish verbs (tr.)</th>
<th>Yiddish verbs (intr.)</th>
<th>Coding in Yiddish</th>
<th>Coding in German</th>
<th>Coding in Russian</th>
</tr>
</thead>
<tbody>
<tr>
<td>boil</td>
<td>kokhn, zidn, tsokkhn</td>
<td>kokhn, zidn, kokhn zikh</td>
<td>labile/anticaus/caus</td>
<td>labile</td>
<td>cau</td>
</tr>
<tr>
<td>freeze</td>
<td>farfrirn</td>
<td>farfrorn vern, frirn</td>
<td>anticaus/caus</td>
<td>labile</td>
<td>equip</td>
</tr>
<tr>
<td>dry</td>
<td>trikenen, oystrikenen</td>
<td>trikenen zikh, oystrikenen zikh, fartriknt vern, trikenen</td>
<td>anticaus/labile</td>
<td>labile</td>
<td>equip</td>
</tr>
<tr>
<td>wake up</td>
<td>ufvekn</td>
<td>ufkhapn zikh, ufvekn zikh</td>
<td>equip/ anticaus</td>
<td>equip</td>
<td>equip</td>
</tr>
<tr>
<td>go out/ put out (fire)</td>
<td>farleshn</td>
<td>farloshn vern, farloshn zikh</td>
<td>anticaus</td>
<td>labile</td>
<td>equip</td>
</tr>
<tr>
<td>sink</td>
<td>ziken</td>
<td>ziken</td>
<td>labile</td>
<td>equip</td>
<td>equip</td>
</tr>
<tr>
<td>melt</td>
<td>tseshmeltn</td>
<td>shmeltzn</td>
<td>cau</td>
<td>labile</td>
<td>anticaus</td>
</tr>
<tr>
<td>stop</td>
<td>opsheltzn</td>
<td>opsheltzn</td>
<td>anticaus</td>
<td>labile</td>
<td>anticaus</td>
</tr>
<tr>
<td>turn</td>
<td>dreyen</td>
<td>dreyen zikh</td>
<td>anticaus</td>
<td>labile</td>
<td>anticaus</td>
</tr>
<tr>
<td>burn</td>
<td>farbrenen, brenen</td>
<td>brenen</td>
<td>cau/labile</td>
<td>labile</td>
<td>equip</td>
</tr>
<tr>
<td>fill</td>
<td>onfill</td>
<td>onfill zikh</td>
<td>anticaus</td>
<td>anticaus</td>
<td>anticaus</td>
</tr>
<tr>
<td>rise/raise</td>
<td>ufheybn</td>
<td>ufheybn zikh</td>
<td>anticaus</td>
<td>anticaus</td>
<td>anticaus</td>
</tr>
<tr>
<td>improve</td>
<td>farbesern</td>
<td>farbesern zikh</td>
<td>anticaus</td>
<td>anticaus</td>
<td>anticaus</td>
</tr>
<tr>
<td>rock</td>
<td>vign</td>
<td>vign zikh</td>
<td>anticaus</td>
<td>labile/anticaus</td>
<td>anticaus</td>
</tr>
<tr>
<td>connect</td>
<td>farbindn</td>
<td>farbindn zikh</td>
<td>anticaus</td>
<td>anticaus</td>
<td>anticaus</td>
</tr>
<tr>
<td>gather</td>
<td>zamen</td>
<td>zamen zikh</td>
<td>anticaus</td>
<td>anticaus</td>
<td>anticaus</td>
</tr>
<tr>
<td>open</td>
<td>efenen</td>
<td>efenen zikh</td>
<td>anticaus</td>
<td>anticaus</td>
<td>anticaus</td>
</tr>
<tr>
<td>break</td>
<td>brekhn</td>
<td>brekhn zikh</td>
<td>anticaus</td>
<td>labile</td>
<td>anticaus</td>
</tr>
<tr>
<td>close</td>
<td>farmakhn</td>
<td>farmakhn zikh</td>
<td>anticaus</td>
<td>anticaus</td>
<td>anticaus</td>
</tr>
<tr>
<td>split</td>
<td>tseshpaltz</td>
<td>tseshpaltz zikh</td>
<td>anticaus</td>
<td>anticaus</td>
<td>anticaus</td>
</tr>
</tbody>
</table>

The equipollent vowel alternations are lost in Yiddish, turning to labile ones (ziken ‘sink (tr., intr.), cf. G. senken vs. sinken), which are being replaced by anticausatives: ufvekn ‘to wake up (tr.)’ vs. zikh ufvekn ‘to wake up’ (cf. G. aufwachen vs. aufwecken). Labile verbs also develop themselves in this direction: farleshn ‘to put
out (fire)' vs. farleshn vern ‘to go out (fire)’ (G. erlöschen ‘<both>’). The absence of the labile model in Slavic languages might support this change.

An interesting strategy of distinguishing between the meanings of a labile verb is the use of inseparable perfectivizing prefixes far-, tse- with the causative counterparts (tseeschmeltsn ‘melt (tr.)’ vs. schmeltsn ‘to melt (intr., 'tr.). The fact that the same prefixes are also used as verbalizers supports their new function as causative markers.

In our sample, less spontaneous verbs, marked by detransitivizer zikh) perfectly correspond to the theory of frequency. But the verbs which describe spontaneous coming to a state show enormous variation, should it be between coding types or competing strategies within the anticausative type (trockenen zikh = getrokt vern = trockenen = ‘to dry (intr.’). Aspectual differences between them will be discussed in the talk.

References


Resources
CMY – Corpus of Modern Yiddish: http://web-corpora.net/YNCSearch/
A Usage-Based Study of the Effects of Typological Distance and L2 Proficiency on L1 Influence During L2 Acquisition
Itamar Shatz (Tel Aviv University)

A learner’s interlanguage is their developing second language (L2) knowledge; its structure depends on a variety of factors, among which are the typological distance between the learner’s native language (L1) and their target L2 (Ellis, 2008; Lakshmanan & Selinker, 2001; Odlin, 2003; Selinker, 1972, 2013). Prior studies show that this is primarily due to a crosslinguistic influence from the L1, which is generally attributed to negative and positive transfer of structures from the L1 to the L2 (Ellis, 2008; Jarvis & Pavlenko, 2008; Odlin, 2003). However, these studies focused mostly on narrow samples, in terms of the number of L1s and the range of L2 proficiency levels which were examined (e.g. Bhela, 1999; Darus & Ching, 2009; Sönmez & Griffiths, 2015). As such, there is a call for a large-scale study on the topic, which could validate prior findings and explore new facets of the effects that typological distance and L2 proficiency have on L1 influence during L2 acquisition.

The present study examines a learner corpus consisting of over 133,000 texts, composed by nearly 38,000 learners of English as a foreign language, who were part of an international English-learning program. These learners came from diverse linguistic backgrounds, which include seven different L1s, and nearly all L2 proficiency levels. The study examines the usage-based error patterns of six linguistics structures in learners’ L2 writing, which relate to verb tense, word order, articles, possession, agreement, and plurality (a subset of agreement). For each structure, the typological distance between learners’ L1 and the target L2 was quantified based on related feature values available in the World Atlas of Language Structures (Dryer & Haspelmath, 2013). The error patterns were based on a standardized annotation of learners’ texts by their teachers, and included a quantification of the number and type of errors that they make for every word in the text, as well as each error’s proportion out of all errors.

The results reveal several key insights. There was significant variation between the different structures. In the case of articles and plurality, an increase in typological distance was associated with increased interference from the L1. Conversely, in the case of agreement and word order, an increase in typological distance decreased L1 interference. In the case of possession and verb tense, acquisition was unaffected by distance. Overall, the findings suggest that typological distance can have varied effects on L1 influence during the acquisition of different structures, as it can either facilitate L2 acquisition, hinder it, or not affect it at all. In terms of L2 proficiency, differences between the speakers of different L1s due to L1 influence remained consistent over time, and even increased in the case of articles.
References


A Big-data Investigation of Natural Language Use Following Terror Attacks
Almog Simchon and Michael Gilead (Ben-Gurion University of the Negev)

Individuals in the western world describe the threat of terrorism as one of the greatest concerns in their daily lives (Chapman University, 2016). As such, it is important to gain a better understanding of the psychological reactions to media reports of terrorist events. One possibility is that in response to the mortal threat that terrorism poses, individuals resort to instinctual and fixed defensive behaviors (Kenrick et al., 2010). However, it is also possible that humans can rely on their intellect and their uniquely-human capacities to carefully reason and find strategic solutions to the danger (Eippert et al., 2007; Miller, 1987). Supporting the latter, Cohn, Mehl and Pennebaker (2004) analyzed a large set of blog posts before and after the 9/11 attacks, and saw increased use of cognitive-analytical language. Cohn et al. interpreted these findings as reflecting individuals’ increased cognitive and intellectual efforts to understand the threatening event. While it may be the case that humans try to rely on higher-order cognition to make sense of horrible acts of terrorism right after it strikes, it is also possible that the stress associated with such events will impair the ability to reason and engage in careful deliberation (Mahoney, Dalby, & King, 1998). Considering these possibilities, in the current investigation we examined how reports of terrorist events affect individuals’ levels of analytical thinking, as reflected in natural language use. We extracted more than 2.5 million messages posted on the social network Twitter in the time before and after seven different terror attacks in the Unites States and the United Kingdom. To measure changes in analytical thinking, we relied on a validated, syntax-based measure of analytical thinking (Pennebaker et al., 2014), implemented in the Linguistic Inquiry and Word Count program (LIWC; Pennebaker et al., 2015). Our analysis relied on a time period beginning with the moment the term “terror” showed an increasing trend in our data until the end of the same day; control segments were exactly one week prior to the terror attack (with the exception of one incident), binned by the same time-based parameters taken from the experimental segments. A Random-effects meta-analysis was employed to compare the level of analytical thinking following the seven different terror incidents to control days, providing evidence for the validity of the effect. Whereas prior investigations that relied on an analysis of semantic features and suggested increased cognitive processing following terror attacks (Cohen et al., 2014), our results qualify these findings by providing evidence that individuals’ language use reflects lower levels of analytical thinking in the hours following a terror attack.
References


Figures

Figure 1. Analytical Thinking and Anxiety plotted in z scores. The upper panel refers to control days, the lower panel refers to experimental days. The attack occurred on June 3rd 2017, around 22:00.

Figure 2. Random-effects meta-analysis comparing the level of analytical thinking following the different terror incidents to control days.