Yiddish passive constructions: a case study based on the new Corpus of Modern Yiddish

Sandra Birzer

1. Introduction

Card files give testimony of what linguistic data retrieval looked like before the age of corpus linguistics: innumerous texts had to be read through in search for a certain expression; matches were excerpted on file cards, and one could only hope that not too many instances had been omitted in the process of reading. If one wished to conduct research on yet another expression, the whole procedure started anew.

The rise of text corpora changed the situation. Generally speaking, a text corpus is a large collection of digitized and annotated texts. The annotation of the texts comprises information about the author and the text on the one hand, and various linguistic information on the other hand, i.e. the part-of-speech-status and morphology of every word is described. The last fifteen years have seen a significant increase of corpus building and linguistic research with the help of text corpora. Working with a text corpus provides linguistic research with profound language data that makes possible even statistical verification of linguistic hypotheses. Recently, some languages received so-called National Corpora, which serve as a modern powerful tool for various linguistic studies and references. Examples are the British National Corpus (http://www.natcorp.ox.ac.uk/, 100 million words), the Czech National Corpus (http://ucnk.ff.cuni.cz/, 500 million words), the Russian National Corpus (www.ruscorpora.ru, 150 million words), and the Eastern Armenian National Corpus (www.eanc.net, 90 million words). All these text corpora can be searched via a freely accessible online interface.

The Corpus of Modern Yiddish (CMY), which is currently being compiled by scholars from Regensburg and Moscow, will contain two subcorpora. The first subcorpus, a balanced one, will contain 10 million word forms from a large variety of texts and will cover the period from 1850 up to today. The second subcorpus will contain contemporary newspaper texts only and will be large enough to make possible large-scale statistical studies. Data retrieval is automatic via an internet interface and

therefore will be available to any interested researcher. In order to make the Corpus of Modern Yiddish accessible to as many interested users as possible, the corpus will be available in both Hebrew characters and YIVO-transliteration in Latin characters. In these respects, the Corpus of Modern Yiddish differs from other resources such as the National Yiddish Book Center (http://www.yiddishbookcenter.org), the Internet Archive (www.archive.org), the collection *Jiddische Drucke* ('Yiddish prints') of the Frankfurt University Library (www.literatur-des-judentums.de) and the numerous websites of Yiddish newspapers and literature, whose texts can be searched only manually.

In the following, we will give a short outline of the selection of texts for the corpus. Then, a small study on passive voice, based on queries to the Yiddish toy corpus already compiled by our team[1], will demonstrate the possibilities of corpus research.

2. Selection of texts[2]

2.1 General remarks on the selection of texts for corpora

The ideal corpus is a balanced multi-purpose corpus that represents a given language in the way this language is used in everyday life: it contains written and spoken language of different genres (for this and the following, cf. Meyer 2002: section 2). Written texts are divided into printed and non-printed ones. Printed texts one may roughly distinguish into

- fiction of various genres
- drama
- memoirs and biographies
- journalism and literary criticism
- scientific, popular scientific and teaching texts
- religious and philosophical texts
- technical texts
- business and juridical texts

To the non-printed texts belong roughly:

- private letters
- business letters
- essays from school or university examinations etc.
- diaries

Spoken texts can be classified as dialogues or monologues, which are delivered spontaneously (e.g. friends talking) or primed (e.g. chairman's address to the club members).

The character of texts from the internet often oscillates between the above-mentioned three major categories. They are all written texts, but blogs and chats are rather informal in style and especially the latter resemble spoken rather than written language.

An ideally balanced corpus is controlled for such sociolinguistic variables as gender, age and educational level of the author[3], dialect variation as well as the social contexts and relationships the text samples were produced in.

Since in our everyday life the percentage of spoken texts we are exposed to is much higher than the percentage of written texts (cf. Meyer 2002: 56), an ideally balanced corpus should contain more spoken than written text. However, hardly any of the large corpora come up to this standard, since the inclusion of spoken texts poses several difficulties: first of all, one must find speakers willing to be recorded; records must not be impaired by the observer's paradox or by insufficiencies of the technical apparatus (cf. Meyer 2002: 56-61), and, finally, spoken texts have to be brought into some kind of transcription that reflects a minimum of oral discourse related phenomena (fillers, self-repairs etc) which is a very time-consuming and expensive process.

Corpora covering large time spans provide special problems (cf. Meyer 2002: 37-38). Ideally, they should provide not only subcorpora balanced within themselves, but also balanced against one another in terms of size. However, this is hardly ever realizable for several reasons: first and foremost, for technical reasons no recordings of spoken language are available for earlier periods, which makes it impossible to include spoken samples into the respective subcorpora. Secondly, the more distant the period is, the scarcer are written texts available for it. This means that when compiling subcorpora of early periods, sociolinguistic variables are necessarily disregarded, largely or completely. For the same reason, one usually has to accept that subcorpora of earlier periods are smaller in size.
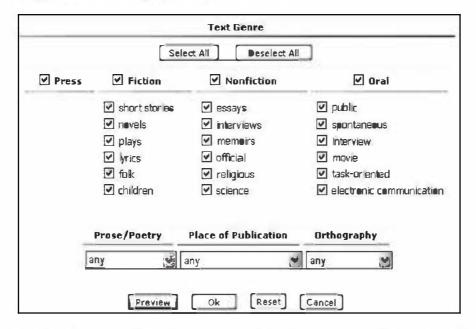
2.2 Selection of texts for the Corpus of Modern Yiddish

The CMY is intended as a multi-purpose corpus; therefore we have decided to make as many search options as possible available to users of CMY. Consequently, we do not offer subcorpora of our own, but users may

specify their own subcorpora with the help of sociolinguistic or time variables (cf. Diagram 1), which is made possible through metatextual annotation of texts.

At the internal level, however, we assume three subcorpora: the first one covers the period 1851–1900, the second one stretches from 1901–1939, and the third one extends over the period 1940-today.[4] Several factors to be discussed below have led to the following allocation of word forms to the above-mentioned periods:

- 2 million word forms for the period 1850–1900
- 6 million word forms for the 1900–1939
- 2 million word forms for the period 1940 – today

Diagram 1. The subcorpus selection form of CMY

As has been mentioned above, a balanced corpus should contain all possible text genres. Besides, in the case of Yiddish it is also vital to represent all dialects (i.e. *Litvish, Poylish* and *Ukrainish*) and varieties of Yiddish (e.g. North American Yiddish), as well as the different ortho-graphies. Translations from other languages are only included in CMY if the overall text production in the given genre is small (e.g. texts about

natural sciences), because translations are often influenced in syntax and lexicon by the original.

Here, some extralinguistic factors come into play. The first of them is the general availability of texts. This comprises not just the mere existence of texts from all text genres – which is the premise for compiling a balanced (sub-)corpus – but also the question whether the material is well-preserved enough for text processing. For example, books whose paper is already decomposing must not be scanned, as the scanning process damages the paper further, and scans of yellowed or stained pages are very difficult to OCR, since OCR programs cannot cope with such defects.

The period between 1901–1939 may be characterized as the heyday of Yiddish. Literary and journalistic texts are available for all genres (the classic Sholem Rabinovich alias Sholem Aleykhem may serve as example for fictional literature, Celia Dropkin – as example of a female poet; New York-based *Forverts* and Warsaw-based *Haynt* – as examples of newspaper texts). Due to the rather large amount of printed material it is possible to take into account all sociolinguistic variables including geographic ones. All other types of printed texts, e.g. scientific texts, as well as non-printed texts are available as well, however access, especially to non-printed texts, is more restricted. A very high number of texts from this period is available in editions of excellent quality, offered by the *National Yiddish Book Center*. Both sociolinguistic and extralinguistic parameters allow compiling a balanced subcorpus for the period 1901–1939, whereas the periods 1851–1900 and 1940-today conform by far not to all these parameters. Therefore we decided to allocate the bulk of word forms to the period 1901–1939. Another issue is the fact that predominantly texts from this period are used for Yiddish language teaching, so that CMY might also become a powerful resource for language teaching.

The 6 million word forms assigned to the period 1901–1939 have to be allocated to different text genres and various geographic areas at the same time.

Table 1. Allocation of word forms to text genres (period 1901–1939)

fictional texts, including poetry and drama	3 million word forms
newspapers	1.5 million word forms
scientific texts, schoolbooks, functional texts (*Gebrauchstexte*: instruction manuals, promotion materials etc.)	1 million word forms
handbooks, official domain	0.5 million word forms
memoirs, personal narratives and letters, journals and diaries	depending on availability

The consideration of geographic areas secures that the different geographic (and thus dialectal) varieties of Yiddish as well as the cultural centres of the given time are represented.

Table 2. Allocation of word forms to geographic areas (period 1901–1939)

Poland and rest of Eastern Europe	2.25 million word forms
Northern America, including Canada	1.5 million word forms
Southern America	0.75 million word forms
Soviet Union	0.75 million word forms
Western Europe and rest of the world	0.75 million word forms

Texts have to be selected in cross-consideration of the two allocation schemes mentioned.

Another issue of corpus building that is often discussed is the length of individual text samples. If one includes 'lengthy' texts, such as novels, the size of the corpus grows very rapidly, yet one risks that the stylistic idiosyncracies of a small number of authors are represented to the corpus user as a representative sample of the given language. Therefore, we will include a maximum of 50 pages from one text. Newspapers are an exception to the rule, as one has to pick numbers from every volume.

3. A case study: passive voice in the Corpus of Modern Yiddish

3.1 Formation of the passive voice in Yiddish

First of all, we will give a small survey of how the passive is treated in Yiddish grammars and various linguistic papers, then we will address the linguistic riddles the passive still poses.

The prototypical passivization strategy in Yiddish, traditionally called "passive construction", is formed with the past participle of a transitive

verb and the auxiliary *vern* 'to become' (cf. among others, Mark 1978: 284–286). We will refer to this type as analytical passive.

(1) *dos hoyz*   *vert*   *geboyt.*
    DET house-NOM   become-AUX.PRS   build-PTCP.PST
    'The house is being built'

(2) *dos hoyz*   *iz*   *geboyt*
    DET house-NOM   be-AUX.PRS   build-PTCP.PST
    gevorn.
    become-AUX.PTCP.PST.PASS
    'The house was built.'

(3) *dos hoyz*   *vet*   *geboyt*   *vern.*
    DET house-NOM   will   build-PTCP.PST   become-AUX.INF
    'The house will be built'

Another type of passivization is the passive with the reflexive *zikh* (4).

(4) *dos bukh* ...   *ruft*   *zikh*   *"lernbukh far onheyber".*
    DET book-NOM   call-3SG.PRS   REFL   textbook for beginners
    (Forverts, 2008.03.21)
    'The book is called 'Textbook for Beginners'.'

Hall (1967: 130) calls this type "reflexive medio-passive with *zikh*". Applying this term to all constructions formed with *zikh*, Hall overstretches the term medio-passive for the following reason. Reflexives in general are highly polysemic items that serve several functions in the middle domain, among them passivization (for an overview of all phenomena belonging to the middle domain cf. Kazenin 2001: 920–923). The term medio-passive is somewhat fuzzy, as its definitions vary. In one place it is equated to the middle voice in general and defined as a "construction in which an intrinsically transitive verb is construed intransitively with a patient as subject and receives passive interpretations: *This fabric washes easily*." (DGTL 1993: 170). This would include other middle voice phenomena such as anticausatives (5) as well, which is not too surprising: Haspelmath (1987: 9) gives a discussion of the aforementioned terms and points out that both terms *middle* and *medio-passive* are applied to anticausatives. Therefore we will abandon the term *medio-passive* and will distinguish between reflexive passive (4), anticausatives (5) and quasipassives (6),

where abstraction from the actor takes place and a modal meaning of possibility is traceable.

(5) *di   tir   efnt   zikh.*
     DET door-NOM   open-3SG.PRS   REFL
     'The door opens.'

(6) *dos bukh   farkoyft   zikh   shlekht.*
     DET book-NOM   sell-3SG.PRS   REFL   bad-ADV
     'The book sells badly.' (close to 'It's impossible to sell the book.')

As a result, we need salient criteria for distinguishing the reflexive passive from anticausative constructions, both of which are formed with *zikh*.[5] As a basis we take Xrakovskij's (1974) definition of the passive.

Xrakovskij (this and the following 1974: 44–45) defines the passive as a diathesis that is characterized by retention of the verb's valencies (7), but a change in their quality (8). The first change in quality is the promotion of the second semantic actant of the verb to first syntactic argument and the second change is the demotion of the first semantic actanct. If the demoted actant is realized syntactically, it fills the verb's second argument slot as a complement; if not, one of the situation's attributes, i.e. an adverbial, fills the second slot. However, in Xrakovskij's view not all adverbials qualify as second argument: only adverbials of time and space (8) are indicative of passive constructions, whereas adverbials denoting the reaction of the second semantic actant upon the action, are indicative of other diathesis constructions.

(7) *der soykher   farkoyft   dos bukh   in der krom.*
     DET vendor-NOM   sell-3SG.PRS DET book-ACC in the shop
     'The vendor sells the book in the shop.'

(8) *dos bukh   farkoyft   zikh   in der krom.*
     DET book-NOM   sell-3SG.PRS   REFL   in the shop
     'The book is sold in the shop.'

However, recent research has shown that in several languages conflated forms allowing also a passive interpretation, which do not conform to Xrakovskij's second precondition, do exist. Yiddish confronts us with this problem as well, since the passive and resultative conflate in the past: *iz geshribn gevorn* 'was written' (cf. Nath 2009: 183). Zumstein (2010)

shows that analytical passive and reflexive forms have default readings, and she offers parameters for ascribing passive or anticausative / resultative meaning to potentially ambiguous forms on the basis of Russian data. These have to be adapted to Yiddish, as the two languages differ in several morphosyntactic parameters. Most important, Yiddish lacks a fully developed system of aspect, which in Russian restricts the forming of analytical passives to perfect verbs and is therefore pivotal for the distribution of passive and anticausative meaning. For Yiddish we can state that the coexistence of analytical passive and reflexive forms persists throughout the whole tense system, and that there are no descriptions available of how the distribution of passive and anticausative meaning is organized in the various lexico-semantic verb classes (although there seems to be some interconnection, as e.g. for *rateven* 'to rescue' only analytical passive forms, but not a single reflexive passive one could be found in the corpus). Adapting Zumstein's (this and the following cf. Zumstein 2010) parameters to Yiddish, we receive the following hypothesis about the distinction of passive and anticausative meaning in Yiddish: The default reading of analytical passive forms is passive, whereas the default interpretation of reflexive forms is anticausative. These are the parameters for ascribing an interpretation to ambiguous forms:

a) if the form in question is built from a verb whose first actant is prototiypically agentive, the passive interpretation is the most probable;

b) if the actor is realized syntactically as a PP, the form has to be interpreted as passive;

c) if the causer of the situation is human and agentive (Zumstein 2010) and / or the situation is embedded into a sequence of actional situations (Zumstein 2010) the default interpretation of the ambiguous form is passive;

d) if the causer is either unclear or non-human and the situation is embedded into a stative context, the ambiguous form is by default interpreted as anticausative.

Zumstein herself has not organized her parameters into a hierarchy. We cannot solve this task in our paper, as our toy corpus is much too small (and too much focused on newspaper texts so far) to give a thorough account of passive and anticausative constructions in Yiddish. Our aim is merely to give a basic outline of these phenomena and to mark the issues that require further investigation in the future.

The reflexive passive in Yiddish might be due to influence from coterritorial languages. For example Polish (cf. Bartnicka et al. 2004: 355–360) and Russian (cf. Xrakovskij 1974) form part of their passive forms with a reflexive marker, whereas the German forms with the reflexive marker *sich* that are derived from transitive verbs are considered reflexive conversives which are only marginally related to the passive (cf. Zifonun et al. 1997: 1856–1858).

(9)   *Dom*   *się*   *buduje.* (POL)
       house-ACC/NOM   REFL   build-3SG
       'The house is being built.'

(10)   *Dom*   *stroit-sja*   *nemeckoj*   *kompanie j.* (RUS)
       house-NOM   build-3SG-REFL   German-INS   company-INS
       'The house is built by a German company.'

Some words are in order on some other phenomena that have been considered peripheral types of passivization.

Hall (1967: 132) mentions the apparent passive, which is in fact a resultative (11). "[R]ezul'tativom imenuetsja forma, oboznačajuščaja sostojanie predmeta, kotoroe predpolagaet predšestvujuščee dejstvie [Resultative is the name for a form denoting the state of an object, which implies a preceding action]", i.e. the resultative accounts "odnovremenno o sostojanii i o predšestvujuščem emu dejstvii, rezul'tatom kotorogo javilos' èto sostojanie. [at the same time of the state and the preceding action, whose result is the state.]" (Nedjalkov and Jaxontov 1983: 7) In many languages, "[r]ezul'tativ javljaetsja odnim iz značenij kakoj-to drugoj formy, vxodjaščej v sistemu zaloga, vida ili vremeni. [the resultative is one of the meanings of some other form, which is part of the voice, aspect or tense system]" (Nedjalkov and Jaxontov 1983: 16) This is also true for Yiddish, where resultative meaning is expressed by the past participle and the auxiliar verb *zayn*, and its coterritorial languages, where the resultative is also expressed mainly with the help of the past participle (cf. Wiemer and Giger 2005, Giger 2009). As a general rule, only telic verbs can form resultatives, in the case of Yiddish even only transitive telic verbs, since only the construction PAST PARTICIPLE + *zayn* may convey resultative meaning in Yiddish.[6] Resultative constructions can be clearly distinguished from analytical passives, as the Yiddish resultative is formed with PAST PARTICIPLE + *zayn*, and the analytical passive with PAST PARTICIPLE + *vern*.

(11) *dos hoyz*   *iz*   *geboyt*   *in 1928.*
      DET house-NOM   be-PRS   build-PTCP.PST   in 1928
      'The house was built in 1928.'

The impersonal[7] construction with *me(n)* is closely related to the passive, since it conveys roughly the same meaning. Although many grammarians consider it preferable to the passive voice (cf. Mark 1978: 286; for the Soviet grammarians' attitude toward the passive see Nath 2009) we will exclude it from our corpus study, as this syntactic construction is already well-described.

(12) *men*   *farkoyft*   *dos bukh*   *in ot der krom.*
      INDF.PRON   sell-3SG.PRS   DET book-ACC   in this store
      'They sell the book in this store.'

3.2 Corpus study of the passive voice

Due to lack of space, we will address only one issue in our corpus study, namely the character of the second syntactic argument of passive constructions.
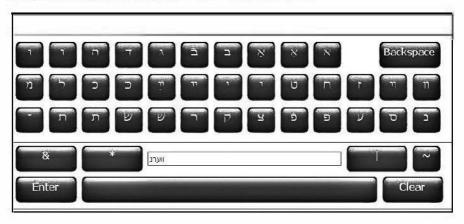
Traditionally it is assumed that the analytical passive splits into so-called 'full passive constructions' (cf. Mark 1978: 287), consisting of *vern* + PTCP.PST + PP representing the demoted agent, and 'semi-passive constructions' (*halbe pasive konstruktsyes*; cf. Shtif 1930: 59), consisting only of *vern* + PTCP.PST. Both Shtif and Mark consider the latter preferable to the former: the 'full passive construction' results in a heavy style (*shvern stil*; Mark 1978: 286), whereas the 'semi-passive construction' can be equated to the impersonal construction with *me(n)* (cf. Shtif 1930: 60-61). This leaves out of sight that the second argument slot of the passive construction can be filled not only by the demoted actor, but by other syntagmatic items as well. The argument structure of the reflexive passive is described even worse. Therefore, subject of our corpus study will be the quantitative distribution of possible second argument types, i.e. we will address the question whether the demoted actor prevails over adverbials as second argument, and whether analytical and reflexive passive show differences in the frequency distribution of these two argument types. Furthermore, we will give a first outline of the features distinguishing passive from anticausative constructions. As the language usage in newspapers is quite specific, we decided to investigate literary texts for

studying the distribution of possible second argument types. Newspaper texts, which at the moment form the larger part of our toy corpus, are consulted only if additional data is required for solving ambiguous cases.

3.2.1 Second argument types of the analytical passive

Since all material for this study has been drawn from our toy corpus, some words on its usage are in order. The corpus allows to search for wordforms and lexemes (in this case all wordforms of the given lexeme are displayed) as well as for grammatical or lexical attributes such as part of speech, case, tense, person etc. The queries for wordforms and lexemes can be entered either in Hebrew characters (diagram 2) or in YIVO transliteration.

Diagram 2. The virtual keyboard for Hebrew characters.

There are also different display options for the search results; probably the most important is the choice between representation in Hebrew characters or transliteration (diagram 3).

Diagram 3. Display options for search results.

Output characters — Yiddish
transliteration
Output layout
Show morphological analysis — yes
Matches per page — 10
Sentences in expanded context — ±1
Sort by

Ok Reset Cancel

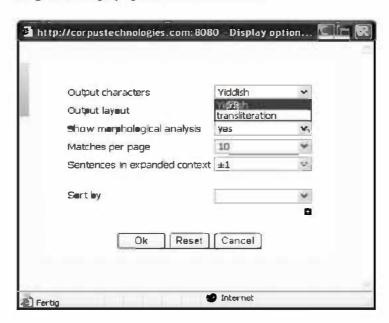Diagram 4 illustrates the search results for one of the queries used for this study (the query is visible on the right side of the diagram). The search results have to be checked for "noise", i.e. wrong matches that result e.g. from homonymy.

Diagram 4. Search results for the query lexeme *vern* 'to become' + past participle. The incorporated Yiddish-English dictionary is activated by moving the mouse to the respective word.

In our toy subcorpus of literary texts (50,000 tokens; the newspaper subcorpus contains about 2.6 million tokens) we found 23 constructions that morphologically qualify as analytical passives. Six of them display a PP denoting the demoted actor (13) as second argument, i.e. they are to be considered as 'full' passives in the traditional terminology. All six PPs denoting the demoted actor are headed by *fun* 'by', yet it must be assumed that the absence of PPs headed by *durkh* 'through' is first of all due to the stylistic features of literary texts and probably also to the small subcorpus size. In newspaper texts PPs headed by *durkh* are rather frequent (14). Another sixteen analytical passive constructions have a complement (15) or an adverbial as second argument (16), thus to be labelled traditionally as 'semi-passives'. Furthermore, one construction without a second argument was found in the corpus (17). We will discuss example (17) below.

(13)   *im hot zikh gedakht,   az   er [...]   vert*
       it seemed to him   that he-NOM   become-AUX.PRS
       geyogt   un   getribn   fun   ot di tayvolim,
       hunt-PTCP.PST   and   chase-PTCP.PST   PREP   DET devils
       vos hobn zikh do farzamlt
       REL had gathered there
       (J. Rosenfeld. *Vos hot pasirt mitn altn?*)
       'It seemed to him that he was hunted and chased by those devils that had gathered there.'

(14)   *... a televizye-program, vos   iz   gefirt*
         a television show   which-NOM   be-AUX.PRS   lead-PTCP.PAST
       gevorn   durkhn   barimtn
       become-AUX.PTCP.PST PREP   famous
       yidish-daytshishn konferansye rozental.
       Jewish-German host Rosenthal
       (*Forverts*, 2007.03.16)
       '... a television show, which was presented by the famous Jewish-German host Rosenthal.'

(15)   *s'ponim   iz   mir   farpatsht*
       DET face-NOM be-AUX.PRS I-DAT   slap-PTCP.PST
       gevorn.
       become-AUX.PTCP.PST
       (I. B. Singer. *Gimpl tam*)
       'I was slapped in the face.'

(16)   ... *hartsraysndike*   *koyles*   *fun*   *an altitshkn,*   *vos*
heartbreaking   cries-ACC   PREP   an old.man-DAT   REL
vert   *shreklekh*   *gepaynikt* ...
become-AUX.PRS   terribly   torture-PTCP.PST
(D. Bergelson. *In a fargrebter shtot*)
'... heartbreaking sounds from an old man who was being tortured
terribly...'

(17)   *dos brikl*   *iz*   *aropgetsoygn*
DET bridge-NOM   be-AUX.PRS   draw_down-PTCP.PST
gevorn .
become-AUX.PTCP.PAST
(L. Shapiro. *Gegesene teg*)
'The small bridge was drawn down.'

Our toy subcorpus is too small to make generalizations, but it is noteworthy
that in literary texts the usage frequency of the full passive construction
amounts to about one quarter of all usages.

One of the grammarians' main objections to the passive construction is
the heavy style it creates. Mark discusses the possibility of passive
constructions drawing on the example (18) and its passive counterpart (19).
Both sentences are characterized by heavy modification both on NP (*oyf
yener farzamlung* 'at this meeting') and sentence level (*nokh a lengerer
debate* 'after a lengthy debate'), which adds clumsiness to the sentence.

(18)   *di merhayt*   *oyf yener farzamlung*   *hot*
DET majority-NOM   at this meeting   have-AUX.PRS
ongenumen   *di rezolutsye*   *nokh a lengerer debate.*
accept-PTCP.PST   DET resolution-ACC   after a lengthy debate
(Mark 1978: 286)
'The majority at this meeting accepted the resolution after a lengthy
debate.'

(19)   *Di rezolutsye*   *iz*   *nokh a lengerer debate*
DET resolution-NOM   be-AUX.PRS   after a lengthy debate
ongenumen   *gevorn*   *durkh*   *der merhayt*
accept-PTCP.PST   become-AUX.PTCP.PST   PREP   DET majority-ACC
oyf yener farzamlung.
at this meeting
(Mark 1978: 286)

'After a lengthy discussion, the resolution was accepted by the majority at this meeting.'

Yet in the passive constructions from the corpus, no matter whether 'full' (13) or 'semi-passive' (16), only the verb is modified if there is some modification at all. If the modification consists of an adverbial phrase or sentence, it is moved to sentence-final position (20). Therefore, the actual sentence structure of analytical passive constructions disproves the apprehensions mentioned above.

(20) *... mayn vuntsh   iz   derfilt*
      my wish-NOM   be-AUX.PRS   fulfil-PTCP.PST
      gevorn   shneler vi ikh hob zikh gerikht.
      become-AUX.PTCP.PST   faster than I expected
      (L. Shapiro. *Gegesene teg*)
      'My wish was fulfilled faster than I expected.'

The verbs *dermonen* und *dertseyln* deserve special attention, as their 'semi-passive' constructions seem to serve exclusively as means of reference to antecedent issues and are used only in past tense (21–22). In the active voice, where the verbs *dermonen* and *dertseyln* do not serve as means of text reference, any tense may be used and syntactic realization of (at least) the first two semantic actants is obligatory (23–25).

(21) *aropkrikhn funem oyvn   iz dem altn,   vi es*
      to crawl down from the stove was for the old man   as it-NOM.EXPL
      iz   shoyn   dermont   gevorn,
      be-AUX.PRS   already   mention-PTCP.PST   become-AUX.PTCP.PST
      nisht azoy laykht geven [...]
      not that easy
      (J. Rosenfeld. *Vos hot pasirt mitn altn?*)
      'Crawling down from the stove, as had been mentioned before, was not that easy for the old man.'

(22) *in eyner a nakht, vos balangt take tsu der zelber tsayt,   vegn*
      in one night   which belongs also to the same time   about
      velkher   es   iz   oybn   dertseylt
      REFL   it-NOM.EXPL be-AUX.PRS   above   tell-PTCP.PST
      gevorn,
      become-AUX.PTCP.PST

hot zikh der alter oyfgekhapt fun shlof mit an umruik gemit ...
the old man startled from sleep with an uneasy mind
(J. Rosenfeld. *Vos hot pasirt mitn altn?*)
'One night, which was part of that very time which has been told about above, the old man startled from sleep with an uneasy mind.'

(23) [...] *er   dertseylt   dem emes* [...]
he-NOM   tell-3SG.PRS   DET truth-ACC
(I. Manger. *Di mayses fun Hershl Zumervint*)
'He tells the truth.'

(24) *kh'hob   ir   alts   dertseylt*
I-NOM:have-AUX.PRS   her   all   tell-PTCP.PST
(I. B. Singer. *Gimpl tam*)
'I told her everything.'

(25) *ober ven ale zenen gezesn baym tish,*
but when all were sitting at.the table
hot   zi [...] *genumen   dertseyln*
have-AUX.PRS   she-NOM   AUX.INCH.PTCP.PST   tell-INF
vegn der levaye [...]
about the funeral
(I. Manger. *Di mayses fun Hershl Zumervint*)
'But when all were sitting at the table, she started to tell about the funeral.'

What still needs to be discussed is the instance of an analytical passive without a second syntactic argument. The form *iz aropgetsoygn gevorn* 'was pulled down' (cf. 17) is ambiguous, as passive and resultative conflate in it. As the wider context is relevant for determining the default reading of an ambiguous form, we will now have a look at it (26).

(26) *der fayf fun parokhod   hot   ibergerisn ir kol.*
DET whistle-NOM   have-AUX   interrupt-PST DET
dos brikl   iz aropgetsoygn   gevorn,
bridge-NOM   AUX draw_down–PTCP.PST   AUX
di pare   hot geton a kokh-aroys
DET steam-NOM   smoulder-3SG.PST.SEMELF
ergets fun der zayt fun shif — un teykef   hot
have-AUX

> *dos shvartse vaser*   *tsvishn der pristan un dem parokhod*
> DET black water-NOM   between the quay and the steamboat
> *genumen   vern   breyter.*
> AUX.INCH   become-INF   broader
> 'The steamboat's whistle interrupted her voice. The small bridge was drawn down, the steam smouldered out somewhere from the ship's side – and the black water between the quay and the steamboat began to become broader.'

The subjects of the sentences neighbouring on the ambiguous form are all non-human (*whistle, steam, water*) and occur in situations that do not imply agentivity (*interrupt* as an undeliberate action, *smoulder, become broader*). Furthermore, the situations cannot be considered sequential in a strict sense. In fact, *dos brikl iz aropgetsoygn gevorn* is the first clause of a paratactic compound sentence and denotes a situation that serves as background for the action in the second clause. Therefore, the ambiguous form has to be interpreted as resultative.

3.2.2 Second argument types of the reflexive passive

Of the 926 word forms with *zikh* that can be found in the toy subcorpus of literary texts, only 13 are potentially passive. We picked six of them for a thorough analysis. We selected our sample so as to have various second argument types (among them absence of the second argument) and a varying degree of agentivity of the first semantic actant.

Our first example (27) has as second argument a PP headed by *fun* and is formed of the verb *oysleshn* 'to extinguish'. *Oysleshn* is a verb whose first semantic actant is not prototypically agentive, as not only animate beings but also natural forces may extinguish fire. This is rather indicative for an anticausative interpretation. On the other side, the PP hints at an animate first semantic actant, namely a bird, which favours a passive interpretation.

(27) *funem fokh fun fligl   hot   zikh*
        by a beat of the wing   become-AUX.PRS   REFL
        dos fayer   oysgeloshn
        DET fire-NOM   extinguish-PTCP.PST
        (J. Rosenfeld. *Vos hot pasirt mitn altn?*)
        'By a wing beat the fire was extinguished.'

Therefore let us check in which other contexts the reflexive construction appears and what is the distribution between analytical and reflexive constructions. All other instances of the reflexive construction appear in contexts, where *oysleshn* is used in a metaphorical meaning (28) and are therefore not helpful. We now turn to the analytical passive constructions. In the present none of the analytical passives is accompanied by a second argument; nonetheless it is obvious that the first semantic actant must be animate (29). For this reason the analytical construction in the present allows only a passive interpretation. The past tense construction *is oysgeloshn gevorn*, which conflates with the resultative construction, seems to be used exclusively with the resultative meaning 'has / had faded' (lit. 'was / were faded') in a rather idiomatic fashion (30). Therefore it is very probable that the reflexive construction with the PP headed by *fun* as second argument is deliberately chosen to convey passive meaning, as the analytical form already bears a rather idiomatic meaning. However, the scarcity of materials allows for no generalisations.

(28)　*sutskever iz shoyn nishto,*　　*s'hot*
　　　Sutskever is already not here　it-EXPL.NOM:have-AUX.PRS
　　　zikh　oysgeloshn　　　　*der letster yidisher shtern ...*
　　　REFL　extinguish-PTCP.PST　DET last Yiddish star-NOM
　　　(*Forverts*, 2006-2010)
　　　'Sutskever is no longer with us, the last Yiddish star has faded.'

(29)　*...punkt　bay nakht　vert*　　　　　　*oysgeloshn*
　　　exactly　by nighttime　become-AUX.PRS　extinguish-PTCP.PST
　　　der elektrisher shtrom ...
　　　DET electric current- NOM
　　　(*Forverts*, 2008.05.30)
　　　'Exactly by the beginning of nighttime the electricity is switched off.'

(30)　*er hot zikh zeyer gebitn,*　　*un　di oygn*　　　*zaynen*
　　　he had changed very much　*and*　DET eyes-NOM　be-AUX.PRS
　　　oysgeloshn　　　　　*gevorn,*
　　　extinguish-PTCP.PST　become-AUX.PTCP.PST
　　　vi bay a sakh alte layt in brayton-bitsh. (*Forverts*, 2007.02.02)
　　　just as with many old people in Brighton Beach
　　　'He had changed very much and his eyes had become dim, just like had happened to many old people in Brighton Beach'

The second example (31) has a PP functioning as adverbial of space and is formed of the verb *kokhn* 'to cook', whose first semantic actant ranks rather high on the agentivity scale. Therefore a passive interpretation is not excluded.

(31)   *in yedn tepl kokht   zikh   boser-khuts,   zayn*
       in every pot boil-3SG.PRS REFL meat-unkosher-NOM   his
       fleysh   kokht   zikh   dortn
       meat-NOM   boil-3SG.PRS   REFL   there
       (D. Bergelson. *In a fargrebter shtot*)
       'In every pot unkosher meat is boiling, his meat is boiling there.'

A glance at more examples shows that the reflexive construction is either embedded into a stative situation (32) or conveys a stative situation serving as background for a person's actions (33). These instances undoubtedly bear anticausative meaning. On the other hand, we found just one example for the analytical construction (34). Although it is used within an idiom, it carries a clearly passive meaning. Given that the wider context (35) of example (31) describes a stative situation, the interpretation of (31) must be anticausative.

(32)   *in mitn hoyf   – etlekhe keslen,   in*
       in the middle of the yard – several cauldrons, PREP
       velkhe   s'kokht   zikh   a shitere zu pekhts
       REL   it-EXPL:cook-3SG.PRS   REFL   a meager soup-NOM
       fun farfoylte burekes
       of rotten beets
       (*Forverts*, 2007.01.12)
       'In the middle of the yard – several cauldrons, in which a meager soup of rotten beets is cooking.'

(33)   *dekt tsu dem top un kokht biz der rayz iz greyt.   beshas   es*
       cover the pot and cook until the rice is ready.   while   it-EXPL.NOM
       kokht   zikh   der rayz,   tseshnaydt   di tsibele.
       cook-3SG.PRS   REFL   DET rice-NOM   cut   the onion
       (*Forverts*, 2007.10.12)
       'Cover the pot and cook it until the rice is ready. While the rice is cooking, cut the onion.'

(34)   *ober keyn zakh vert*   *nit azoy*   *gegesn,*
      but nothing   become-AUX.3SG.PRS   not so   eat-PTCP.PST
      vi   *es*   *vert*   *gekokht.*
      how   it- NOM   become- AUX.3SG.PRS cook-PTCP.PST
      (*Forverts*, 2009.04.17)
      'But nothing is eaten in the state as it is cooked.'

(35)   *iber der shtot shteyt ... di ... zun, un fun ale koymens geyt roykh. in*
      yedn tepl kokht zikh boser-khuts, zayn fleysh kokht zikh dortn. (D.
      Bergelson. *In a fargrebter shtot*)
      'Above the city the sun is standing, and smoke curls from all
      chimneys. In every pot unkosher meat is boiling, his meat is boiling
      there.'

In example (36) were are dealing with the verb *truknen* 'to dry' whose first semantic actant ranges very low on the agentivity scale, as drying is a process that can hardly be controlled. No second argument is encountered here, and the interpretation of the reflexive form is clearly anticausative.

(36)   *s'hot*   *zikh getruknt*   *vesh.*
      it-NOM.EXPL:have-AUX.PRS   REFL dry-PTCP.PST   laundry-NOM
      (I. B. Singer. *Gimpl tam*)
      'Laundry was drying.'

Examples (37-39) will be dealt with in one block, as they are all ambiguous in the sense that they can be interpreted as passives or quasipassives.

(37)   *nor punkt in yener tsayt*   *hobn*   *zikh*   *fun*
      but just at this time   have-AUX.PRS   REFL   from
      foder-shtibl   *derhert*   *a por opgerisene oysgeshreyen*
      the front room   hear-PTCP.PST   some staccato cries-NOM
      (D. Bergelson. *In a fargrebter shtot*)
      'But just at this time from the front room could be heard some
      staccato cries.'

(38)   *baym aynforhoyz zet   zikh   der onhoyb*
       at the inn   see-3SG.PRS   REFL   DET beginning-NOM
       fun a goyisher gas.
       of a non-Jewish street
       (L. Shapiro. *Gegesene teg*)
       'From the inn the beginning of a non-Jewish street can be seen.'

(39)   *un take nor mit dem   derklert   zikh   es,*
       and as well only with that   explain-3SG.PRS   REFL   it-NOM.EXPL
       vos ... iz der alter gezesn ... oyfn oyvn ...
       what the old man was sitting on the stove
       (J. Rosenfeld. *Vos hot pasirt mitn altn?*)
       'And also only with this fact can be explained why the old man was
       sitting on the stove.'

Contrasting examples (37–39) with analytical forms (40–42), differences
become obvious: the analytical forms, no matter whether they display a
second syntactic argument or not, all imply an agentive first semantic
actant. In (40) this agentive actant is the art connoisseur attending the
session, in (41) the Yiddish community and in (42) the author of the
explanation that follows. They can be interpreted only as passives.

(40)   *leybu loyns   kunst   muz   nisht nor   gehert*
       Leybu Loyn's   art-NOM   must-3SG.PRS   not only   hear-PTCP.PST
       vern,   nor oykh   gezen   vern.
       become-AUX.INF   only as well   see-PTCP.PST   become-AUX.INF
       (*Forverts*, 2007.02.09)
       'Leybu Loyn's art has not only to be heard, but to be seen as well.'

(41)   *di poylishe* [geshtaltn]   *vern ...*   *gezen*
       DET Polish figures-NOM   become-AUX.3SG-PRS   see-PTCP.PST
       mit   di oygn funem yidishn tsiber. (*Forverts*, 2008.11.07)
       PREP   DET eyes of a Yiddish community
       'The Polish figures are contemplated with the eyes of a Yiddish
       community.'

(42)   *Ober ... inem tsveytn punkt ... vert*   *bald*
       but in the second paragraph   become-AUX.3SG.PRS   soon
       derklert,
       explain-PTCP.PST

az yisroel iz dos heymland fun di araber-palestiner ...
that Israel is the homeland of the Arab Palestinians
(*Forverts*, 2007.06.29)
'But in the second paragraph it is soon explained that Israel is the
homeland of the Arab Palestinians.'

Example (43) shows another reflexive construction with the "classical"
prerequisite for the quasipassive interpretation, namely negation. This is
one more piece of evidence that the reflexive constructions listed above are
to be interpreted as quasipassives. In this context it should be mentioned
that Russian, one of the languages Yiddish was in permanent contact with,
has the lexemes *slyšno* 'to be audible', *vidno* 'to be visible' and the
reflexive form *ob-jasnjaetsja* 'can be explained' that also imply
quasipassive meaning. It is quite probable that Yiddish formed its
quasipassives according to this model.

(43) *der sholem*   *zet*   *zikh*   *ober*   *nisht*
      DET peace-NOM   see-3SG.PRS   REFL   but   not
      oyfn horizont *oykh nokh di finf yor* ...
      on the horizon   also after those five years
      (*Forverts*, 2007.04.06)
      'However peace is not visible on the horizon even after those five
      years.'

3.4 Conclusion

In Yiddish there exist two constructions allowing for a passive inter-
pretation, namely the analytical passive formed of the auxiliary *vern* 'to
become' and the past participle, and the reflexive construction with *zikh*.
The default reading of the analytical construction is the passive, whereas
the default interpretation of the reflexive passive is the anticausative.
Therefore, the analytical passive construction is prototypical and the
reflexive construction is rather marginal for conveying passive meaning.
    The existence of default readings implies that both forms allow other
interpretations as well. In the past tense the analytical passive conflates
with the resultative; the resulting ambiguity is resolved as follows. The
analytical passive construction is characterized by the promotion of the
verb's second semantic actant to subject of the passive construction; the
actor is demoted.

Analytical passive constructions are embedded into a sequential context and the verbs they are formed of have a human, preferably agentive first semantic actant. In the ideal case the demoted actor is syntactically realized as PP headed by *fun* 'of' or *durkh* 'through'. Non-sequential contexts and verbs with a non-human first actant trigger the resultative interpretation of the analytical form PAST PARTICIPLE + *vern*.

The default reading of non-lexicalized reflexive forms, on the other hand, is the anticausative one. There is reason to assume that reflexive constructions acquire passive meaning if they are formed of verbs with a highly agentive first actant which is syntactically realized as PP headed by *fun*. Unfortunately, the data at our hands is too scarce to prove this hypothesis, but our study may serve as an outline of the research questions concerning the passive, anticausative and quasipassive which still have to solved for Yiddish.

Abbreviations

SEMELF – semelfaktive
INCH – inchoative
EXPL – expletive

Note

*   The Corpus of Modern Yiddish is a joint venture of scholars from Regensburg and Moscow supported by the Fritz-Thyssen-Stiftung (grant number 10.09.1.065) and the Binyumen Shekhter Foundation for the Advancement of Yiddish. We also wish to thank the National Yiddish Book Center for permitting usage of their electronic library, the *Forverts* for permitting usage of the texts on their website, Raphael Finkel for providing his Yiddish OCR software and Roland Meyer for counselling.

1.   By the time of writing this article, the toy corpus contained roughly 2.65 million tokens and was accessible only to members of the team compiling the Corpus of Modern Yiddish.

2.   This chapter contains the work results of all team members: Sandra Birzer, Mikhail Daniel, Holger Nath, Evita Wiecki, Björn Hansen, Vladimir Plungian, Mikhail Kudinov, Aleksey Polyakov, Nikita Bezrukov, Alexandra Polyan, Elena Luchina.

3.   Here and in the following, the term *author* refers to both writers and speakers.

4. The 1740–1850 period will be covered by another corpus, which will be the result of the project "Historische Syntax des Jiddischen mit transkribiertem Textkorpus zum älteren Jiddisch (HJS) [Historical Syntax of Yiddish with a transcribed text corpus of Older Yiddish]" (http://www.indogermanistik.uni-jena.de/index.php?auswahl=135&ident=LE), hosted at the Friedrich-Schiller-Universität Jena., Germany. This made it possible for our project to concentrate on the period from 1850 onwards. Furthermore, it is generally assumed that the period of Old Yiddish literature lasted until 1780, then followed by Hasidic and Haskalah literature in the period of 1780 – 1890, whereas 1864 is usually considered to be the begin of Modern Yiddish literature. From a linguistic point of view, the years 1740–1850 constitute an early period of Modern Yiddish that is characterized by transitional processes and relatively meagre literary output. Literature of this time is mainly of a religious nature. Only few legal, administrative or technical texts, as well as personal writing, come from that time. Yiddish journalism did not yet exist, and there are practically no women authors. Spoken texts are also unavailable. Therefore, it would not have been possible to build a balanced subcorpus of the 1740–1850 period.

5. Due to reasons of space, we have to refrain from including a thorough study of the quasipassive.

6. The Slavonic languages may express resultative meaning also with the help of the perfective adverbial participle denoting anteriority:

   usevšis'   *za stol,*   *gosti*   *načali*   *est'*.
   take_place-CONV behind the table   guests-NOM   start-PL.PST   eat-INF
   'Having taken place at the table, the guests started to eat.'

   Since the Yiddish gerundive, the functional equivalent of the Slavonic adverbial participle, conveys only simultaneity (cf. Mark 1978: 335), Yiddish cannot express resultative meaning with the help of the gerundive.

7. Note that this is the traditional term; in other grammarian traditions sentences of this type are referred to as indeterminate personal ones.

References

Bartnicka, Barbara., Björn Hansen, Wojtek Klemm, Volkmar Lehmann, and Halina Satkiewicz
     2004   *Grammatik des Polnischen.* (Slavolinguistica 5) München: Otto Sagner.

DGTL – Trask, Robert Lawrence
     1993   *A Dictionary of Grammatical Terms in Linguistics.* London – New York: Routledge.

Giger, Markus
     2009   Der Resultativ in den slavischen Sprachen. In: Berger Tilman, Karl Gutschmidt, Sebastian Kempgen, and Peter Kosta (eds.), *Die*

slavischen Sprachen, 269–274. (Handbücher zur Sprach- und Kommunikationswissenschaft 32) Berlin: de Gruyter.

Hall, Richard M. R.
  1967   Yiddish syntax: phrase structure rules and optional singulary transformations of the modern standard language. PhD thesis. New York University.

Haspelmath, Martin
  1987   *Transitivity Alternations of the Anticausative Type*. (Arbeitspapiere Nr. 5) Cologne: Institut für Sprachwissenschaft der Universität zu Köln.

Kazenin, Konstantin I.
  2001   Verbal reflexives and the middle voice. In: Haspelmath Martin, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible (eds.), *Sprachtypologie und sprachliche Universalien. Ein internationales Handbuch. Vol.* 2., 916–927. Berlin – New York: Walter de Gruyter.

Mark, Yudel
  1978   *Gramatik fun der yidisher klal-shprakh*. [Grammar of the Yiddish Standard Language.] New York: Congress for Jewish Culture.

Meyer, Charles F.
  2002   *English Corpus Linguistics. An Introduction*. Cambridge University Press.

Nath, Holger
  2009   The Passive in Soviet Yiddish. *Groninger Arbeiten zur Germanistischen Linguistik* 49: 182-199.

Nedjalkov, Vladimir P., and Sergej E. Jaxontov
  1983   Tipologija rezul'tativnyx konstrukcij. [Typology of resultative constructions] In: Vladimir P. Nedjalkov (ed.), *Tipologija rezul'tativnyx konstrukcij. (rezul'tativ, stativ, passiv, perfekt)* [Typology of resultative constructions. (resultative, stative, passive, perfective)], 5–41. Leningrad: Nauka.

Shtif, Nokhem
  1930   *Yidishe stilistik*. [Yiddish stylistics.] Kharkov: Tsentraler farlag far di felker fun F.S.R.R.

Van Valin, Robert
  2005   *Exploring the Syntax-Semantics Interface*. Cambridge.

Wiemer, Björn, and Markus Giger
  2005   *Resultativa in den nordslavischen und baltischen Sprachen. Bestandsaufnahme unter arealen und grammatikalisierungstheoretischen Gesichtspunkten*. München: Lincom Europa.

Xrakovskij, Viktor S.
  1974   Passivnye konstrukcii. [Passive constructions.] In: Aleksandr A. Xolodovič (ed.) *Tipologija passivnyx konstrukcij. Diatezy i zalogi.* [Typology of passive constructions. Diatheses and genera verbi.], 5–45. Leningrad: Nauka.

Zifonun, Gisela, Ludger Hoffmann, and Bruno Strecker
    1997   *Grammatik der deutschen Sprache*. Band 3. Berlin – New York: Walter de Gruyter.
Zumstein, Mirjam
    2010   Abgrenzung von Passiv und Rezessiv (Antikausativ) im Russischen. *Zeitschrift für Slavische Philologie* 67/ 1, 163–184.