Topicality, Predicate Prototypes, and Conceptual Space

Patrick Murphy
Duke University

Abstract

This paper has the goal of investigating the nature of membership within the category 'passive' and cross-linguistic comparison of constructions, 'passive' and otherwise. Topicality measures were collected from the Uppsala Corpus of Russian for passive and active uses of the Russian verbs *pisat'/napisat'* 'to write,' *davat'/dat'* 'to give,' and *zabyvat'/zabyt'* 'to forget.' Croft's (2001) notion of plotting constructions in 'conceptual space' is exploited as a means of cross-linguistic comparison using these topicality measures. Examining the conceptual space of various voice constructions with these Russian verbs, Croft's generalizations are upheld, their position being consistent whether Referential Distance or Topic Persistence is used as a measure. Finally, data from other typological discourse studies is plotted, noting where various voice constructions pattern, and how this data fits into Croft's model.

Keywords: passive, voice, prototypes, topicality, radical construction grammar, conceptual space, corpus linguistics

1. Introduction

Passive constructions are perhaps the most widely studied grammatical phenomena within generative grammar. In passive voice constructions, the subject is the logical object of the action denoted by the verb. In English, sentence (1a) is in the active voice, while (1b) is in the passive voice.

(1) a. Pat kissed Kim.
   b. Kim was kissed (by Pat).

Both sentences in (1) refer to the same event and the same actors. The difference is that the object of the verb *kiss* in (1a), *Kim*, is the subject of (1b). *Pat*, the subject of the active verb *kiss* in (1a), is either left out entirely of (1b), or is optionally expressed obliquely by means of a *by*-phrase (indicated by parentheses). The purpose of the construction seems to be to focus attention on the logical object (*Kim*), or to de-emphasize the logical subject (*Pat*). The passive voice in English is marked by use of a passive participle (*e.g., kissed*) periphrastically with a form of *be*, although passives are also formed periphrastically in English with *get*:

(2) a. Kim got kissed.
   b. Kim got kissed by Pat.

How much of this characterization of English passives can be applied to other languages? Other languages have constructions with similar structural or functional characteristics; for this reason, such constructions are typically labeled as "passives." Examples from Persian, Dutch, Russian and Irish are given below.

(3) a. Ahmed   košté   šod.
      Ahmed   killed   become

'Ahmed was killed.'
  b. Er   wordt   door   de   jongens   gefloten.
     There   became   by   the young men   whistled
     'There was some whistling by the young men.'
  c. Vchera   ego   ubili.
     Yesterday   him   killed-3PL
     'Yesterday (they) killed him.' = 'Yesterday he was killed.'
  d. Bhí Seán   á   bhualadh   aici.
     Was John   to-his   hit-NOM   at-her
     'John was being hit by her.'

The Persian passive in (3a) is very similar to its English passive equivalent: *Ahmed was killed*. The choice of the auxiliary verb used to form the passive periphrastically is different (*become* instead of *be*), but it is easy to see why both the Persian sentence in (3a) and the English sentence in (1b) would be considered "passives."

The Dutch verb in (3b) is morphologically marked as a passive participle; that is, the verb is formed differently from the active form, and in the same way that the verb would be in the Dutch translation of *Kim was kissed*. But the verb is an intransitive one; the literal English translation of (3b)—*There was whistled by the young men*—is impossible. The Dutch construction in (3b) is called an "impersonal passive" because it has passive morphology and the logical subject is expressed as an oblique. But unlike the English passive, this construction cannot serve to focus attention on the logical object; there is no logical object in this sentence.

The Russian example in (3c) is morphologically active; Russian has passive participles which are morphologically distinct from active forms. The verb *ubit'* 'kill' here is ex-pressed as an active form, conjugated to agree with an absent third person plural subject. In some contexts, this absent subject is generic, and is meant to be interpreted as "someone." Like the English passive, the function of this construction seems to be to de-emphasize the logical subject of

the sentence. Unlike the English passive, the predicate is morphologically active. Is this an example of a passive?

Finally, the Irish construction in (3d) shares several features with the English passive. The passive is formed periphrastically with a form of *be*, and the logical subject is expressed as an oblique. Yet unlike the English passive, the Irish passive in (3d) is associated with contexts in which the purpose is to **emphasize** the logical subject. This is exactly the opposite of the functional use of the English passive. How do we reconcile these facts? This paper has the goal of investigating the nature of membership within the category "passive" and cross-linguistic comparison of constructions, "passive" and otherwise.

2. Conceptual Space

Croft (2001) takes issue with the many revisions to syntactic theories that inevitably follow from the examination of grammatical phenomena like the passive. Croft's book *Radical Construction Grammar* argues that most of the theory and methodology behind current syntactic analyses of passives is fundamentally flawed. His argumentation is firmly rooted in his typological approach. Based on the diverse set of structural characteristics of the constructions he examines, Croft states that theories seeking to find universal structural categories cannot hope to succeed.

As may be gathered from the name of the book, Croft's theory is essentially one of construction grammar, albeit one that differs from "mainstream" construction grammar. Croft argues that constructions are primitives of syntactic representation, and they deter-mine grammatical categories, which are not primitives. Similarly, syntactic relations are limited to those between a construction and any elements of which it is composed.

Croft decides that there are no universal grammatical categories,

and that all constructions are language-specific. Despite this, the author does not dismiss the search for linguistic universals; instead, he contends that these must be quantified in terms of form-to-meaning mappings.

Presenting many different voice alternations from a diverse selection of languages, Croft argues that constructions are language specific and "that constructions as cross-linguistically valid configurations of morphosyntactic properties do not exist" (Croft 2001:283).

Croft asserts that what is needed to successfully characterize a cross-linguistic grammatical construction is to specify what part of conceptual space will be examined, and to place a language's use of a construction within it. In discussing active, passive, and inverse constructions, Croft focuses on the conceptual space of a two-participant event, V.
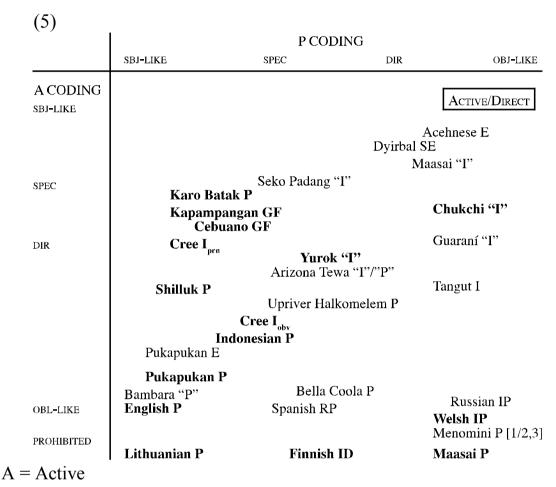
(4)

$$
\begin{array}{cccc}
 & \text{P:} & & \\
 & 1, & 2 & < & 3 \\
\text{A:} & 1, & - & \\
 & 2 & & - \\
 & < 3 & & \\
\end{array}
$$

(Croft 2001:284)

This space maps the status of two referentially disjoint speech act participants, the A and P, where these denote the usual syntactic roles (as per Myhill (1992), A is the "subject" of a transitive verb and P the "object"). The numbers indicate the participants in the event: 1 = the speaker, 2 = the addressee, and 3 = some other participant. The <sign indicates that the only ordering here is partial; the addressee is ordered before "other" participants, but the order is undefined for speaker and addressee. This is a speech act participant (SAP)

hierarchy. Croft does not in this work address voice in constructions with more than two participants, nor does he examine middles, reflexives, or antipassives.

He then proceeds to present data from a number of languages, examining different voice constructions and how they map onto the space in (4), based on person/number restrictions on their arguments. Without reproducing all of his examples here, the aim is to show that constructions labeled "active" or "passive" in these languages do not occupy the same conceptual areas on the author's map, nor do they share structural properties. They differ as to whether A is coded as an Oblique or not, whether P is coded like a Subject or not, and whether V is morphologically distinct from the Active V or not (Croft 2001:295). This, he argues, reinforces his position that there are no universal characteristics of passive constructions. Moreover, he says the same of all voice types.

Croft advocates comparing structural properties of voice constructions cross-linguistically by examining the coding of A and P, which can be defined as Subject-like, Object-like and Oblique-like with respect to the basic voice construction. These properties define a multidimensional "syntactic space," upon which language-particular constructions can be plotted (Croft 2001:312). His representation of the syntactic space for voice constructions appears in (5):

(5)

	P CODING			
A CODING	SBJ-LIKE	SPEC	DIR	OBJ-LIKE
SBJ-LIKE				**Active/Direct**
				Acehnese E
			Dyirbal SE	
				Maasai "I"
SPEC		Seko Padang "I"		
	Karo Batak P			
	Kapampangan GF			**Chukchi "I"**
	Cebuano GF			
DIR	**Cree I$_{prn}$**			Guaraní "I"
		Yurok "I"		
		Arizona Tewa "I"/"P"		
	Shilluk P			Tangut I
		Upriver Halkomelem P		
	Cree I$_{obv}$			
	Indonesian P			
	Pukapukan E			
OBL-LIKE	**Pukapukan P**			
	Bambara "P"	Bella Coola P		Russian IP
	English P	Spanish RP		**Welsh IP**
PROHIBITED				Menomini P [1/2,3]
	Lithuanian P	**Finnish ID**		**Maasai P**

A = Active

boldface: verb form distinct from Active/Direct verb form

P = Passive

IP = Impersonal Passive

RP = Reflexive Passive

ID = Indefinite

Scaling (A top to bottom, P left to right):

I = Inverse   A case: sbj < erg < dir < obl < prohibited

E = Ergative

A agr: sbj < nonsbj < special < none < prohibited

SE = Split Ergative   P case: sbj < dir < obj

GF = Philippine Goal Focus P agr: sbj < special < obj/none

(Croft 2001:313)

The upper right corner of this diagram is meant to represent the basic (transitive) voice type of a language, and various language-specific constructions are compared by plotting them against this basic voice type. Croft then offers the following universal:

(6) If there is a contrast between a basic and non-basic voice…then the semantic map of the basic voice will include the upper right corner of the conceptual space in [(4)], and the semantic map of the nonbasic voice will include the lower left corner of the conceptual space in [(4)].

(Croft 2001:315)

As this universal is formulated on the conceptual space (4), which itself was based on the SAP hierarchy, Croft can offer the following corollary generalizations:

(7) If there are two or more voice constructions for two-participant Vs in a language, then the construction in which A most clearly outranks P in the SAP hierarchy—i.e., 1, 2, 3—is the basic voice construction. If there is an alternation between two constructions which satisfy the first condition, then the most frequent construction will be–or become–the basic voice construction.

(Croft 2001:315)

The author notes that it is "widely accepted" that "the SAP hierarchy is a conventionalized stand-in for topicality," given that the speaker and addressee are most topical to a conversation (Croft 2001:315). Thus he amends his syntactic space to map onto A and P topicality/salience, and expands it to cover both transitive and intransitive types as (8):

(8)

SALIENT ⟷ ABSENT

A:

SALIENT

ABSENT

P:

"unergative"

antipassive

active/direct

inverse

passive

anticausative
("unaccusative")

▭ (bold)	situation types discussed in Croft's chapter
▭	intransitive (one-participant) situation type
⬭ (dotted)	transitive (two-participant) situation type

(Croft 2001:317)

Croft's approach is important because it offers us a way to compare passive constructions cross-linguistically, despite these constructions' lack of shared structural characteristics. I will use his conceptual space in the comparison of passive constructions in section 4.

3. Russian Passive Data and its Collection

One of the criticisms typically leveled at linguists is that the data we use is too idealized or superficial in nature—that our theories

break down when they are applied to the actual utterances that comprise the speech of any given language.

So I seek to temper what linguists think we know about passives by examining data from a specific language in more detail. Moreover, I want to look at the "messy" data produced naturally by speakers and see what parts of the universal story are compatible or incompatible with the language-specific data. I have chosen Russian as the language to examine in this way.

In this section, I will describe the various features of Russian constructions which are labeled "passives." I will describe data collected for three Russian verbs from the online Uppsala Corpus of Russian, and the methodology behind my data collection.

3.1. Passives in Russian: Forms and Functions

Russian has several different constructions which could be labeled passive—and are so labeled in traditional grammar books—and I will discuss each in turn before making any judgments on the appropriateness of that categorization.

3.1.1. Imperfective Passives

The Russian verbal system has a rich aspectual system, and nearly all verbs are morphologically marked as either perfective or imperfective in aspect.

Imperfective verbs form passives in the following way:

(9) The Russian Imperfective Passive
    i.   The logical object is inflected for Nominative case.
    ii. The verb appears with the reflexive particle *-sia*.
    iii. The logical subject may optionally appear, but does so in the Instrumental case.

The following are examples of imperfective passive verbs, with (10a) and without (10b) expressed agents:

(10) a. Dom   stroit-sia   (rabochimi).
      house.NOM   build.IMPERF-REFL   workers.INST
      'The house is being buile (by the workers).'
   b. Dom   stroit-sia.
      house.NOM   build.IMPERF-REFL
      'The house is being built.'

Note that Russian has no copula in the present tense.

3.1.2. Perfective Passives

Perfective verbs in Russian form passives in the following way:

(11) The Russian Perfective Passive
   i.   The logical object is inflected for Nominative case.
   ii. A passive participle appears with an inflected form of *byt'* 'be.'
   iii. The logical subject may optionally appear, but does so in the Instrumental case.

The passive participles which appear in perfective passives are typically "short form" ad-jectives, which have truncated versions of the standard adjectival endings. The following are examples of perfective passives, with (12a) and without (12b) expressed agents:

(12) a. Dom   byl   postroen   (rabochimi).
      house.NOM   was   build.PERF   worker.INST
      'The house was built (by the workers).'
   b. Dom   postroen.
      house.NOM   was build.PERF

'The house was built.'

Interestingly, when conducting searches on the Internet for the participial forms, I discovered instances of a different construction which is not found in Russian grammars. Judging from the authors' names and email addresses and the construction's relative productivity, it seems to be a natural one for at least some native speakers of Russian. In this construction the agent/logical subject is expressed by a possessive construction in Russian, formed by the preposition *u* and a noun phrase in the Genitive case. Examples include:

(13) a. Krome Gosudarstvennogo, u   neë
     besides state   at   her.GEN
     napisany   eshchë   chetyre   gimna
     write.PERF   another   four   hymns
     'Besides the National [Anthem], she has written four other anthems.'
  b. U   menia   dan   pochti   doslovnyi eë
     at   me.GEN   give.PERF almost   verbatim her/its
     perevod.
     translation
     'I have given her/its translation almost verbatim'

Since in Russian, the *u* + GEN construction is used to express possession, *i.e.,* 'have,' it seems that native Russian speakers may employ a perfective passive which more closely resembles English *I have X-en* Y. Clancy (2000) points to a few similar examples of possessive phrases marking "a possessor of a passive event," but only a couple of the examples he presents—from Allen (1978) of "dialectal Russian" —are passives of the type discussed here.

Another strategy employed by Russian speakers is to encode the logical subject – typically expressed by an NP in the Instrumental

case – as a possessive modifying the logical object.

(14) Lennonovskaia "Come Together"   byla   napisana
Lenon's.NOM   come together   was   write.PERF
kak   gimn   ego   predvybornoi   kampanii.
like   anthem   his.GEN   campaign.GEN   campaign.GEN
'Lennon's "Come Together" was written as if to be the anthem of his election campaign.'

Here, [John] *Lennon* is clearly the logical subject of the matrix clause, as he is the author of the song "Come Together." Rather than expressing this fact by encoding *Lennon* as an Instrumental, it is expressed as an adjective modifying the logical object. Likewise, examples can be found where logical subject NPs in Genitive case modify the logical object:

(15) Kontsert   Borisa Tishchenko   napisannyi pochti
concert.NOM   Boris Tishchenko.GEN   written   almost
tridtsat'   let   nazad.
thirty   years   ago
'Boris Tishchenko's concert was written almost thirty years ago.'

Whether this is a generally available strategy or one necessitated by other factors is a question I will return to later.

Russian can also have passives with sentential objects, though by count they are far less common than lexical NP objects:

(16) Na   ix   zamknutyx   litsax   bylo   napisano
on   their   closed   faces   was   written
chto   oni   oskorbleny   moim   schast'em.
that   they   insulted   my   happiness

'It was written on their closed faces that they were insulted by my happiness.'

Here there is no lexical NP for the passive participle *napisano* 'written' to agree with in case, so the participle is expressed in the neuter case.

3.1.3. Other Passives

There are other constructions which Russian grammars term "passive," although the works surveyed earlier differ as to whether they are truly passives or not. I will return later to the question of their status, and simply discuss their characteristics here.

***Oni*-passives:** Russian has forms which are functionally passive, but morphologically active. The verb is an active form, marked for third person plural agreement, but with a maximally non-topical subject or generic or unknown interpretation.

(17) Dom   postroili.
     house.NOM   build.3PL
     'They (unspecified) built the house.' = 'The house was built.'

These "passive" forms do, however, imply human involvement, whereas the past passive participle can lack this implication.

Russian has a rich system of verbal adjectives, or participles, which have the following characteristics:

(18) Russian Participles
     i.   Can be of either past or present tense.
     ii. Can be of either active or passive orientation.
     iii. Perfective verbs only form past participles.
     iv. With few exceptions, passive participles can be formed

from transitive verbs only.[1]
    v. Combine the syntactic and morphological properties of verbs and adjectives.

<div align="right">(Launer 1974:102)</div>

These participles appear both in the "short" forms of the perfective passives above (12), and in longer forms with full adjectival morphology. They can appear with agent phrases in the Instrumental case.

(19) Propisnye   istiny   zabytye   chelovekom.
      essential.NOM truths.NOM forgotten.NOM man.INST
      'Essential truths, forgotten by man'

Russian also has an impersonal construction which is often labeled a "passive" because the Agent/logical subject is expressed as an oblique, and the translation in English is usually that of a passive:

(20) Volnami   uneslo   lodku.
      waves.INST carry away.3SG NEUT   boat.ACC
      'The boat was carried away by the waves.'

However, the similarity ends here, since the logical object is coded as an object, and the verb is always in third-person singular neuter form. Additionally, there is a semantic restriction that only allows Agents which are natural forces.

---

[1] Exceptions include forms of *dostignut'* 'achieve,' *rukovodit'* 'manage/run,' and *upravliat'* 'govern/control' (Launer 1974:102).

3.2. Russian Corpus Data

Studies like Keenan (1985) provide typological generalizations about the relative marked-ness of passive constructions. It would be useful to further bolster these arguments with examples from a large body of data from a specific language. To this end, I examined the occurrence of passive verbs within the Uppsala Corpus of Russian, queried via the Internet from the University of Tübingen's web site.[2] This web site describes the corpus as follows:

(21) The Uppsala Corpus (Upsal'skij korpus russkix tekstov) consists of some 600 Russian texts with a total of one million running words (word tokens), equally divided between informative and literary prose. The in-formative texts are from between 1985 and 1989, while the literary texts, whose vocabulary does not date as quickly, cover a longer period, 1960-88. The corpus does not include poetry or drama. Within the given framework, considerable effort has been made to en-sure as representative and varied a corpus as possible. The informative texts are drawn from 25 different subject areas: economics, foreign affairs/foreign policy, ideology/domestic policy, party matters, Soviet society, social issues, defence, education, law, history, culture, linguistics, medicine/health care, psychology, environment/ ecology, agriculture, engineering, information technology, space research, energy, biology, geology/geography, physics, chemistry and sport. Certain areas which were felt to be more important are represented by a larger volume of texts. The literary half of the corpus comprises work by the following 40 authors: Abramov, Ajtmatov, Astaf'ev, Baklanov, Bek, Belov, Bitov, Bondarev, Dubov, Ganin,

---

[2] http://www.sfb441.uni-tuebingen.de/b1/en/korpora.html

Gladyshev, Granin, Grekova, Goncharov, Iskan-der, Kaverin, Kazakov, Kochnev, Kozhevnikova, Nagibin, Lixanov, Lidin, Paustovskij, Pogodin, Pristavkin, Troepol'skij, Rasputin, Shcherbakova, Simonov, Solouxin, Shmelev, Tendrjakov, Tokareva, Tolstaja, Trifonov, Vasil'ev, Vorobl'ev, Zalygin and Zorin. Here, too, there is unequal representation, with a larger amount of writing by the better-known authors.

(Gattnar 2002)

The corpus is also described in greater detail in Lönngren (1993).

While online corpora such as the Uppsala Corpus afford wonderful access to primary data, it is true that the use of the Corpus must be tempered by the limitations of both corpora and online resources.[3] Therefore, as I encountered various complications, I had to make decisions regarding my methodology. In this section, I seek to clarify and motivate those decisions I made regarding my data collection.

In a pilot study, I decided to investigate an aspect of Russian passives which would complement the data and generalizations presented earlier, and one which could be measured in a reasonably concrete manner: the "topicality" of the logical subjects and logical objects of passive predicates. I identified a set of Russian verbs that I felt would be likely to occur frequently as passives within the corpus, and ran queries on the appearance of these verb forms. My initial list of predicates included those listed in (22):

---

[3] I have participated in preliminary discussions on the responsible use of Internet resources with other linguists. The linguists who participated in these discussions were Hyug Ahn, Ashley Batten, Biljana Belamaric-Wilsey, Jamie Bishop, Sung-ho Choi, Dagmar Divjak, Sean Flanagan, Laura A. Janda, Anne S. Keown, Patrick M. Murphy, James Phillips, and Jennie Powers. Our preliminary suggestions for methodology involving the Internet in linguistic inquiry are summarized as Janda (2003), but clearly more thoughtful discussion among an even broader cross-section of linguists is needed.

(22) <u>Verb</u>   <u>Meaning</u>
      davat'/dat'   'give'
      pisat'/napisat'   'write'
      zabyvat'/zabyt   'forget'

The pairs listed are imperfective and perfective forms, respectively.

To gauge both the scope of this undertaking and get a feel for the data, I chose to start my analysis with a pilot test of a small number of examples of *pisat'/napisat'* 'to write.' Given Russian's rich inflectional and derivational systems, and in keeping with my earlier presentation of Russian passives, I identified the following forms as 'passive:'

(23) <u>Verb</u>   <u>Word Form</u>

pisat'/napisat'	Present Passive Participle	Masc.	n/a
		Fem.	n/a
		Neut.	n/a
		Plur.	n/a
	Past Passive Participle	Masc.	*napisannyi*
		Fem.	*napisannaia*
		Neut.	*napisannoe*
		Plur.	*napisannye*
	Imperfective (reflexive + INST)	Masc. Pres.	*pishetsia*
		Fem. Pres.	*pishetsia*
		Neut. Pres.	*pishetsia*
		Plur. Pres.	*pishutsia*
		Masc. Past	*pisalsia*
		Fem. Past	*pisalas'*
		Neut. Past	*pisalos'*

		Plur. Past	*pisalis'*
Perfective	Masc.	*(byl) napisan*
byt' 'to be' +	Fem.	*(byla) napisana*
short form adj.		
	Neut.	*(bylo) napisano*
	Plur.	*(byly) napisany*
oni ('they')- form w/o subject	Imperf. Pres.	*pishut*
	Perf. Pres.	*napishut*
	Imperf. Past	*pisali*
	Perf. Past	*napisali*

Thus for *pisat'/napisat'*, 18 distinct morphological forms would require 18 searches of the corpora.[4]

These searches would nevertheless turn up some data which would need to be discarded. For instance, recall that the *oni*-forms function as passives only when there is no overt subject; *i.e.,* they are functionally passive, but not morphologically distinguished from active verbs. Therefore many of the "hits" returned when searching for the *oni*-forms were ignored: specifically, those *oni*-forms with overt subjects, or those with non-overt subjects whose referents were clearly identified within the context of the passage in which they were found. Since the "passive" use of *oni*-forms is one in which verbs are inflected for third person plurals but which refer to no one in particular, we omit the results of hits in which the third person plural

---

[4]   Being adjectives, each of the past participle forms also could appear in Nominative, Accusative, Gen-itive, Prepositional, Instrumental or Dative case, with differing declension by number and gender. This means besides those Nominative forms listed on the table, the forms *napisannogo*, *napisannom*, *napisan-nomu*, *napisannym*, *napisannuiu*, *napisannoi*, *napisannykh*, and *napisannymi* would all be indicative this past passive participle. To simplify this table, I have omitted them in this count.

subject has an antecedent in the relevant passage.[5]

Similarly, our earlier description of imperfective passive verbs showed that they are formed by using reflexive morphology and a logical subject in the Instrumental case. We ignore other verb forms with such reflexive morphology which lack logical subjects in the Instrumental case; these would not be passives and thus not pertinent to our present inquiry.

For each of these forms, I ran a search of all the Uppsala corpora, and asked for a context of 2000 characters, which I thought likely to cover the immediate context of their appearance sufficiently to measure things like Referential Distance and Topic Persistence (RD, (28) and TP, (29)). The results of these searches were then analyzed in detail.

For each occurrence of the verb form in question, I noted any overt subjects or obliques, as well as the RD and TP of the logical objects and subjects, as defined in (28) and (29). This involved translation that was at times difficult, as I only had 2000 characters of context rather than the whole of the book or article, and was coming in in the middle of a conversation, as it were. Still, the 2000 character limit seemed sufficient for the collection of pragmatic data.

3.2.1. Interviews and Conversations

The Uppsala Corpus contains many interviews from periodicals like *Ogonëk*. The contexts in question include multiple-participant

---

[5] It is still possible that such an antecedent exists, just earlier in the text than the section captured during the corpus search. However, it is usually clear from context whether or not the "they" subject of a verb is a generic "they" that is unknown or one that refers to something that has been identified and talked about. At any rate, the topicality of such an antecedent would have to be maximal, since its Referential Distance would extend beyond the scope of the corpus segment in the same way that a new discourse item would. Conversely, the Topic Persistence of the item would be zero, and thus also of minimal topicality.

dialogues. The RDs and TPs for various NPs would be different depending on whether one counted back (or forward) over *all* clauses of the conversation, regardless of participant, or only over those clauses uttered by a single participant. That is, is topicality about how recently or often a referent is mentioned by the speaker, or about how recently or often it is mentioned by anyone participating in the conversation? My feeling is that the natural way to think about topicality is the latter way. Therefore, I counted clauses in interviews no differently than I did for texts that were produced by a single author.

But corpora generally provide other problematic cases that are often overlooked by linguists. Consider the following exchange from one of the works of fiction in the corpus:[6]

> (24) a. Telefon   davno   zvonil, no   Rudakov   dolgo
>   telephone far away rang   but Rudakov   long
>   dumal,   chto   zvenit   u   nego   v
>   thought   that   rang   at   him   in
>   ushakh, poka   ne   dogadalsia sniat'   trubku.
>   ears   until have sense   to remove   receiver
>   'A telephone rang far away, but Rudakov thought for a long time that there was ringing in his ears, until he had the sense to pick up the receiver.'
>   b. "Da,"   skazal   muchitel'no,   "da, da."
>   yes   said   agonizingly   yes yes
>   "Yes," he said agonizingly, "yes, yes."
>   c. "Andrei!   Eto   Sof'ia   Alekseevna."
>   Andrei   this   Sof'ia   Alekseevna
>   "Andrei! This is Sof'ia Alekseevna."

---

[6] From "Identifikatsiia Zhenshchiny," by I. Tarasevich, *Ogonëk*, No. 26, 24 June 1996.

d. "Da," povtoril Rudakov.
   yes repeated Rudakov
"Yes," repeated Rudakov.

It is relatively straightforward to count matrix clauses and ignore relative clauses, but how does one classify an utterance like *"yes, yes"*? For the sake of consistency – and perhaps simplicity – I chose to count such utterances as clauses for both measurements.[7]

Also, following Myhill (1992), I considered the maximum value for RD to be 20, and did not count relative clauses, but did count other types of subordinate clauses. Likewise, Myhill describes two "standard" methods for counting Topic Persistence: counting the number of consecutive following clauses in which the referent of the NP being coded is mentioned, or counting how many of the following 10 clauses contain references to the referent of an NP. Given the possibility of somewhat fragmentary dialogue, and a pragmatic limit on the contexts of my searches of 2000 characters, I opted to measure TP by the second standard rather than the first.

Russian might be considered a pro-drop language, in that pronouns are often optional when the context and verbal inflection make clear the referent, meaning the NPs we are searching for are not always overt. Still, they are entailed by verbal inflection.

(25) a. Ivan prishël ko mne.
       Ivan come.3SG.PAST to me.DAT
       'Ivan came over to my place.'
   b. Skazal chto khotel stat'
      Say.3SG.PAST that want.3SG.PAST become.INF
      aspirantom.
      graduate student.INST

---

[7] That is to say, I counted "Yes!" as a clause, and "Yes, yes." as *one* clause. Repeated clauses were counted as one, though there were relatively few examples of these.

'He said he wanted to become a graduate student.'

Assuming that (25a) and (25b) are two contiguous sentences in a text, it is clear from the inflection of *skazat'* 'to say' that *Ivan* is the subject of the matrix clause of (25b). In the corpus data, this happens often, and in such cases, I would count verbs whose inflection entails a referent as a substantiation of that referent, when lacking an overt subject. For example if the two sentences in (25) were the only two of a corpus, I would consider the TP of *Ivan* to be at least 1, since *skazal* 'he wrote' entails its subject, *Ivan*.

The complementizer phrase *chto khotel stat' aspirantom* is a bit more problematic, as here *khotel* could have as its subject either *Ivan* or some other appropriate third-person singular referent from earlier in the discourse. Naturally, in such cases one has to examine the context to determine whether or not it was Ivan or someone else who was exhibiting masochistic tendencies.

3.2.2. Practical Limitations

While the Uppsala Corpus affords unprecedented access to primary Russian data, it is a vast corpus, and as it provides no easy way to search for some of the information I was after, another issue that arose was how much searching was feasible without becoming bogged down in an overload of data.

Consider a search for relevant data for the verb *pisat'/napisat'* 'to write.' With search contexts of 200 characters, these were the number of pages of text produced for the various verb forms:

(26)

Verb Form			Pages Produced
Past Pass. Part.	Masc.	*napisannyi*	67
	Fem.	*napisannaia*	79
	Neut.	*napisannoe*	144
	Plur.	*napisannye*	139
Imperfective	Masc. Pres.	*pishetsia*	147
	Fem. Pres.	*pishetsia*	same
	Neut. Pres.	*pishetsia*	same
	Plur. Pres.	*pishutsia*	50
	Masc. Past	*pisalsia*	26
	Fem. Past	*pisalas'*	24
	Neut. Past	*pisalos'*	57
	Plur. Past	*pisalis'*	31
Perfective	Masc.	*(byl) napisan*	197
	Fem.	*(byla) napisana*	261
	Neut.	*(bylo) napisano*	1140
	Plur.	*(byly) napisany*	220
oni-form	Imperf. Pres.	*pishut*	879
	Perf. Pres.	*napishut*	81
	Imperf. Past	*pisali*	739
	Perf. Past	*napisali*	416

Undoubtedly, many of these pages would contain data outside the parameters described here; for example, it is likely that the majority of the *pisali* 'they wrote' forms do not have a generic third person plural subject, characteristic of the passive. Still, the only way to discern this is to wade through the corpora search output. Also, this table again ignores the various declensions of the adjectival participles.

Multiplying this by my earlier list of verbs (22) leads to a great

number of pages, and the comparison to active forms could generate a prohibitive sample size. Additionally, examining each context is a labor-intensive exercise, as it involves jumping into the middle of a discourse and manually searching for RD and TP evidence.

I felt a balance needed to be struck here between thoroughness and attention to other issues which might be deemed most important. To that end, I constrained myself to the first 100 occurrences of each verb form, at least when the number of occurrences within the corpus exceeded 100. This would seem to produce a sample which is manageable but nonetheless sufficiently representative and statistically significant.

Similarly, I limited the number of verbs which were thoroughly examined in this way. Looking at the first 100 occurrences of each of the verb's passive forms was constrained to searches of *pisat'/ napisat'* 'to write,' *davat'/dat'* 'to give,' and *zabyvat'/zabyt'* 'to forget.'

3.2.3. How the Numbers were Derived

I include here some examples from the Russian corpus, demonstrating how I arrived at the RD and TP numbers for each of the constructions examined. In the example below, the original Russian text is presented on the left, while the English translation is provided on the right. First I list the antecedent for the logical object of the passive sentence. The following number, in brackets, represents the number of non-relative clauses intervening between this antecedent and the sentence containing the passive. Following this, in brackets, is the number of occurrences of the logical object and logical subject over the next ten non-relative clauses.

RUSSIAN

stsenarii . . . [2 clauses]. . . Esli pervye serii byli produmany do melochei, to **dal'she stsenarii** podchas pisalsia na s"emochnoi ploshchadke soobshcha *vsei s"emochnoi gruppoi.* [0 instances of **stsenarii**] [0 instances of *vsei s"emochnoi gruppoi*]
P/logical OBJ RD: 4, TP: 0
A/logical SUBJ RD: 20, TP: 0

ENGLISH

script…[2 clauses]…If the first series were thought through to detail, then **further scripts** were written on the set by *the whole film crew.* [0 instances of **script**] [0 instances of *whole film crew*]

In this example, the antecedent to the logical object is *stsenaria* 'scenario,' which appears in the third clause before the example sentence. There is one non-relative clause at the begin-ning of the sentence before the actual passive clause: *to dal'she stsenarii podchas pisalsia na s"emochnoi ploshchadke soobshcha vsei s"emochnoi gruppoi* 'then further scripts were written on the set by the whole film crew.'

Thus the RD measure for *stsenarii* 'scenario' in this example was 4. Had *stsenarii* been referred to in the clause just before the passive one, it would have had the lowest possible RD, 1. Had *stsenarii* not appeared within 20 clauses previous, the RD would have been the maximum, 20.

The TP measure for *stsenarii* in this example was 0, equal to the number of times *stsenarii* appeared in the 10 clauses immediately following. The logical subject, *vsei s"emochnoi gruppoi* 'whole film crew,' had a RD of 20 and a TP of 0. The film crew was not referred to in the previous 20 clauses, nor in the 10 clauses immediately following.

Here are additional examples of some of the passives I found, and

the RDs and TPs of their arguments:

RUSSIAN
"Come Together, Join the Party"…*Lennonovskaia* **"Come Together"** <u>byla napisana</u> kak gimn ego predvybornoi kampanii. [o instances of **"Come Together"**] [0 instances of *Lennon*]
P/logical OBJ RD: 1, TP: 0
A/logical SUBJ RD: 20, TP: 0

ENGLISH
"Come Together, Join the Party"…*Lennon's* **"Come Together"** <u>was written</u> like the anthem of his election campaign. [0 instances of **"Come Together"**] [0 instances of *Lennon*]

RUSSIAN
Shekspira…[6 clauses]…**Nekotorye original'nye teksty** Shekspira <u>napisany</u> slabo. [1 instance of **ix**, referring to **teksty**; 1 instance of **stixi**, referring to **teksty**] [2 instances of *Shekspir*]
P/logical OBJ RD: 20, TP: 2
A/logical SUBJ RD: 7, TP: 2

ENGLISH
Shakespeare…[6 clauses]…**Several original texts** of *Shakespeare's* <u>are</u> weakly <u>written</u>. [1 instance of **them**, referring to **texts**; 1 instance of **poems**, referring to **texts**] [2 instances of *Shakespeare*]

RUSSIAN
gimn…[12 clauses]…*Nina Vladimirovna*...Krome Gosudarstvennogo, *u neë* <u>napisany</u> **eshchë chetyre gimna**. [5 instances of **gimn**] [1 instance of *vam*, referring to *Nina Vladimirovna*]
P/logical OBJ RD: 14, TP: 5
A/logical SUBJ RD: 1, TP: 1

ENGLISH
Anthem…[12 clauses]…*Nina Vladimirovna* …Besides the National [Anthem], *she* <u>has written</u> **another four anthems**. [5 instances of **hymn**] [1 instance of *you*, referring to *Nina Vladimirovna*]

Again, relative clauses were not counted, following convention. But as I mentioned earlier, there were problematic cases involving interjections and other utterances which might or might not have been considered clauses. I chose to consider these interjections clauses for purposes of my RD and TP counts. Consider (24) again, repeated below as (27):

(27) a. [Telefon   davno   zvonil,]
        telephone   far away   rang
        [no   Rudakov   dolgo   dumal]
         but   Rudakov   long   thought
        [chto,   zvenit   u   nego   v   ushakh,]
         that   rang   at   him   in   ears
        [poka   ne   dogadalsia   sniat'   trubku].
         until   have   sense   to remove   receiver
        'A telephone rang far away but Rudakov thought for a long time that there was ringing in his ears until he had the sense to pick up the receiver.'
    b. "[Da,]"   [skazal   muchitel'no,]   ["da, da]."
         yes   said   agonizingly   yes yes
        "Yes," he said agonizingly, "yes, yes."

    c. "[Andrei]!   [Eto   Sof'ia   Alekseevna]."
        Andrei   this   Sof'ia   Alekseevna
     "Andrei! This is Sof'ia Alekseevna."
    d. "[Da,]" [povtoril   Rudakov].
        yes   repeated   Rudakov
     "Yes," repeated Rudakov.

Note the placement of brackets. These mark what I considered to be the boundaries of clauses for purposes of topicality measures.

4. The Conceptual Space of Voice Constructions

Having collected data on the Referential Distance (RD) and Topic Persistence (TP) of several types of passives within a Russian corpus, I now examine what that data might tell us about their use, and how they compare with the characterization of passives cross-linguistically in terms of participant profiles. First, let us consider the possible complications involved in these measures.

4.1. Measurement and Problems of Measurement

The terms of Referential Distance (28) and Topic Persistence (29) have been used to mea-sure the topicality of NPs since Givón (1983). These terms may be understood as follows:

(28) Referential Distance (RD)
    This measurement assesses the gap between the previous occurrence in the discourse of a referent/topic and its current occurrence in a clause, where it is marked by a particular coding device. The gap is thus expressed in terms of number of clauses to the left. The minimal value that can be assigned

is thus 1 clause, which is maximally continuous.[8]

(29) Topic Persistence (TP)

> We measure persistence in terms of the number of clauses to the right—i.e., in subsequent discourse from the measured clause—in which the topic/participant continues as an uninterrupted presence as a semantic argument of the clause; an argument of whatever role and marked by whatever grammatical means. The minimal value that can be assigned is thus zero, signifying an argument that decays immediately, i.e., of the lowest persistence.
>
> (Givón 1983:15)

While these measure have been used in a consistent fashion via agreed-upon convention, it is not clear that studies relying upon these measurements (such as Cooreman (1984), Cooreman et al. (1984), Cooreman (1988), Rude (1988), Noonan (1990), Myhill (1992), Siewierska (1993), and Xing (1993)) necessarily have results that can be straightforwardly compared. Let us consider some of these potential barriers to comparison.

As I have noted in section 3, due to the relatively small size of my samples, I chose to measure TP as an integer with a floor of 0 and a ceiling of 10, taken as a count of how many times over the next 10 clauses the referent reappeared. This is, following Myhill (1992), an accepted measure of TP:

> "This count . . . is an indicator of **cataphoric** topicality, that is, the importance of an NP in the **following** discourse. There are two types of topic persistence counts. One type counts the number of consecutive following clauses in which the referent of the NP

---

8 By convention, the maximum value for RD is set to 20; if the referent's antecedent in discourse is more than 20 clauses previous, we consider the value to be 20.

being coded is mentioned. . . . The second standard way to count Topic Persistence is by counting how many of the following ten clauses contain references to the referent of an NP. In this case, the clauses do not have to follow consecutively. I am not aware of any study comparing the relative advantages and disadvantages of these two methods."

(Myhill 1992:36)

Givón's definition of TP in (29) is an alternate approach where TP is measured as number of contiguous clauses the referent appears in before it is not referred to again. These are two different, but conventionalized, standards of measure for the same term; we should not expect these numbers to be too similar, and thus it might be difficult to compare them directly.

Likewise, linguists computing these numbers might differ in their actual methods of deriving them. The definitions above have little to say about the exact methods used to compute RD and TP; when we hope to have precise numbers which are statistically significant, vagueness of methodology is undesirable. Again, following Myhill (1992), we know that relative clauses are to be ignored, but there are few other guidelines provided concerning these numbers:

"Established methodology is not to count relative clauses but to count other types of subordinate clauses. I do not know of any explicit justification or argument for this distinction and this is an characteristic of the theory presented here which could be changed by a study designed to evaluate the advantages and disadvantages of different systems if counting."

(Myhill 1992:36)

In fact, I suspect the diverse nature of the type of texts examined can influence methodology and results. I have described in section 3 how the many interviews present in the corpora I used presented

numerous utterances of the type "Yes, yes!" I chose to consider these utterances full clauses for purposes of my counts, but other linguists faced with the same data may not have. Or it could be the case that other corpora would lack such interjections altogether, or have a preponderance of relative clauses which might affect the numbers one way or the other.[9]

Register could have a similar impact. Noonan's (1990) study of Irish actives and passives notes different numbers for spoken and written Irish. It is for such reasons that I have tried to describe my methodology carefully. The corpora used are also readily accessible for examination.

Yet the difficulty goes beyond there being two accepted standards for counting TP, or the largely unavoidable differences in subject matter and linguists' methodologies. It is also unclear whether these numbers will be at all similar across languages, or even whether they should be. Cross-linguistically, we might expect to see a great deal of variation in RDs and TPs, depending on the language. Thus we should not be surprised to see a great deal of variation in RD and TP measures between languages or even within different texts of the same language.

Finally, our discussion of the topicality of passive arguments has been consistently described in relative terms. Whether characterized as a transformation or lexical rule which applies to "de-emphasize the logical subject" or "promote the logical object," the characterizations are couched in prominence relationships between NPs.

For example, Croft (2001) summarizes two analyses of the relationship between topicality and voice—that of Thompson (1994) and Cooreman (1987)—as (30):

---

[9] One area of methodology—or at least reporting—that is of particular concern here is granularity in topicality results. The topicality data I will present in section 4.2 includes both short and long passives. It is not always clear in other studies' topicality data on passives whether authors are reporting only on one type of passive or both.

(30) Two Analyses of Topicality and Voice Constructions
     Thompson (1994:48):   *Cooreman (1987:76):*

Thompson		Cooreman	
A	P		
A↑	P↑		
A	P↑	Active/Direct	AGT > PAT
A↓	P↑	Inverse	AGT < PAT
A↓	P	Passive	AGT ≪ PAT
A↓	P↓		
A	P↓	Antipassive	AGT ≫ PAT
A↑	P↓		
A↑	P		

(Croft 2001:316)

Thompson (1994) characterizes his As and Ps in terms of "increased prominence" or "decreased prominence," indicated by up or down arrows after the letters, respectively. These are measured against the measures for Active/Direct clauses. For Thompson, passive use corresponds to any of three situations: a predicate with a A of decreased topicality and a P of increased topicality, a predicate with a A of decreased topicality and a P of normal topicality, or a predicate with a A of decreased topicality and a P of decreased topicality.

Cooreman's (1987) system matches relative topicality of Agent and Patient against each other to correspond to different voice constructions. The '>' sign indicates "greater topicality than"; the '≫' sign, "much greater topicality than," and so on. These categories are all defined in relative terms, by comparing the topicality of A and P to those in actives and passives, or Agent versus Patient. Cooreman thus characterizes passives as having Agents of much higher topicality than Patients.

While these systems differ as to what the salient topicality features are, the classifications are also imprecise. While Thompson's (1994) classifications are more fine-grained than are Cooreman's in

that there are more situation types presented, neither is set up to capture every nuance of the relationship between the two roles.

Furthermore, there is no reason to think that the choice between actives and passives is deterministic for a language. Rather, we are examining a probabilistic tendency towards the choice of one construction over another based on the relative salience of referents in the discourse. We are suggesting a statistical probability.

For this reason, it is worth noting the possible problems with measurement that might be statistically significant.[10] One potential problem is that of *censored data*. "Censored data refers to a set of observations where some of the values are known only up to a boundary value" (Thode 2002:159). Recall that the ceiling for RD by convention is 20; if the referent's antecedent in discourse is more than 20 clauses previous, we consider the value to be 20. So our data set contains numerous instances of 20. The problem is that these 20's are not really 20's: they could be 20's, or 40's, or 1000's; our data does not tell us the actual referential distance. This means inaccuracies when attempting to test for normalcy, which statisticians might hope to counter by various means, but which I will not. Likewise, the TP has a ceiling of 10 by my measure, but it is certainly possible, and in many cases probable, that the referent will be brought up again in the discourse but past our threshold of 10. In a sense, this number is censored as well. When presenting my data later I will return to these questions, but for now I simply note that RD could be a less accurate measure than TP for this reason, or at least *prima facie* might require more adjustment on our part to improve its accuracy.

4.2. Russian Argument Topicality Data

All things being equal, we might expect that the topicality measures

---

[10]   I am indebted to Chris Wiesen for pointing out the problem of censored data to me, and for talking through the statistical issues of my analysis.

for arguments of Russian verbs to be the same. A wide array of studies on the passive have found that constructions so labeled vary widely in their structural and semantic characteristics cross-linguistically. Should we assume that pragmatic characteristics, particularly those of argument topicality, vary as well? And if there is variation, is it constrained to cross-linguistic variation, or would we find variation among passives within the same language?

One possibility for answering these questions would be to examine the topicality data for a significant number of arguments of verbs within a particular language, and look for variation among those data. This is what I did for three Russian verbs: *pisat'/napisat'* 'to write,' *zabyvat'/zabyt'* 'to forget,' and *davat'/dat'* 'to give.' My methods are described in section 3.

I averaged the RD and TP scores for passives of all three Russian verbs separately, a smaller number of actives, and for all three verbs combined. The results are presented below as (31):

(31) Topicality Measures for Arguments of Russian Passives

		Tokens	P RD	P TP	A RD	A TP
pisat/napisat' **'to write'**	Overall	324	13.12	0.92	18.83	0.22
	"Typical" examples	285	12.69	0.92	18.75	0.23
	Opportunists	39	16.28	0.92	7.72	1.61
	Possessive 'u'	12	16.33	1.42	3.42	2.50
	Actives	20	15.53	0.73	7.50	2.50
zabyvat'/zabyt' **'to forget'**	Overall	316	12.43	1.16	19.19	0.12
	"Typical" examples	313	12.48	1.15	19.19	0.12
	Opportunists	3	7.67	2.67	8.33	2.33
	Possessive 'u'	2	10.50	1.00	12.00	0.50
	Actives	20	11.00	1.67	1.85	2.55
davat'/dat' **'to give'**	Overall	74	15.00	0.65	19.08	0.23
	"Typical" examples	68	15.26	0.65	19.00	0.25

	Opportunists	6	12.00	0.67	9.20	2.00
	Possessive 'u'	5	12.40	0.40	9.20	2.00
	Actives	20	16.70	0.50	9.50	1.50
All three	Overall	714	13.01	1.00	19.01	0.18
verbs						
	"Typical" examples	666	12.85	1.00	18.97	0.18
	Opportunists	48	15.21	1.00	7.93	1.70
	Possessive 'u'	19	14.68	1.11	5.84	2.16
	Actives	60	15.44	0.76	6.28	2.18

Here "typical" examples refers to those passives which were classified as neither opportunists nor possessive passives, while "overall" shows data from all three groups: "typical" examples, opportunists, and possessives. The numbers represent the average Referential Distances and Topic Persistences for the Ps (logical object) and As (logical subject), respectively.

Low RD values correspond to high topicality/salience. For RD, then, "salient" means low RD averages, with antecedents being referred to relatively recently in the text, while high RD averages correspond to "absent" coding. Conversely, high TP averages would correspond to high saliency and low TP averages to low saliency.

We can see a significant difference between the scores in RD and TP for active and passive forms for As. For the actives, As are much more topical than Ps, having substantially lower RDs and higher TPs. Thus active As are referred to more recently in the discourse and continue to be referred to more often than Ps in actives.

Likewise, active Ps tend toward higher RDs and lower TPs than passive Ps. The Ps of active clauses are thus less salient and topical than those in passives, referred to less recently in the discourse and referred to again less often. This reaffirms the pragmatic role of the passive as focusing attention on a P of higher-than-normal topicality.

The difference in topicality between the two arguments of the actives is greater than the difference in topicality between the two

arguments of the passive. This fact suggests that the normal state of affairs for the active is that the A is much more salient than the P. For the passive, the P is more salient than it would be in the corresponding active, and the A is much less salient than in the corresponding active.

5. The Conceptual Space of Russian Passives

Recall that Croft (2001) advanced the idea of comparing structural properties of voice constructions cross-linguistically by examining the coding of the A and the P. These he placed on two opposing axes, defining what he called "conceptual space" and "syntactic space." Although he argues that constructions are language-specific, and that universal structural comparisons are flawed, he suggests that by comparing the coding of As and Ps and plotting them in such space, language-specific constructions can be compared (see (4), (5) and (8)).

Croft placed language-specific constructions in such conceptual and syntactic space by initially appealing to a speech act participant hierarchy, in which A and P were partially ordered or ranked based on their participation in the event (these were 1 = speaker, 2 = addressee, 3 = other, as in (4)). Croft later noted that this hierarchy equates to salience/topicality, stating that it is "widely accepted" that "the SAP hierarchy is a conventionalized stand-in for topicality," given that the speaker and addressee are most topical to a conversation (Croft 2001:315). So his opposing axes in (8) are labeled in terms of A and P topicality, wherein the author places transitive and intransitive situation types.

Yet this description suffers from the same sort of imprecision that Thompson (1994) and Cooreman (1987) do. While Croft emphasizes the role of his conceptual space in allowing the cross-linguistic comparisons that structural comparisons cannot make, to a large

extent his placement of constructions within this space is subjective.[11] In his syntactic space for voice constructions (5), Croft's placement of the Shilluk passive is somewhere between that of the Cree pronominal inverse and the Pukapukan passive on the A coding axis because its special A agreement places it "somewhere between" the marking for Cree and Pukapukan that Croft is using as a guide.
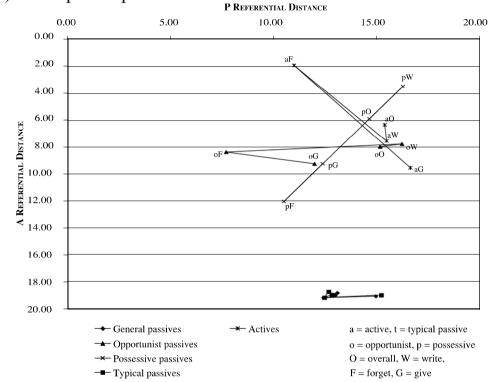
In an attempt to remove some of the imprecision of Croft's categorization system, I turned to the topicality averages I had derived for each verb, and replaced Croft's generic "SUBJECT-LIKE ↔ OBJECT-LIKE" and "SALIENT ↔ ABSENT" descriptive continuums with the topicality measures of RD and TP.

Recall that low RD values correspond to high topicality/salience. For RD, then, "salient" means low RD averages, with antecedents being referred to relatively recently in the text, while high RD averages correspond to "absent" coding. Conversely, high TP averages would correspond to high saliency and low TP averages to low saliency.

In addition to attempting to take away most of the subjectivity from these notions of "salience" and "topicality," I also sought to exploit another notion of Croft's. If constructions from different languages could be compared by means of such a space, could not constructions within the same language?

Substituting RD measures for saliency produces the syntactic space (32):

---

[11] Recall that Croft states that "A syntactic space, defined on structural properties of constructions that are cross-linguistically valid, allows one to develop a proper typological classification of languages that respects the continuum of constructions found in the world's languages for any given region of conceptual space" (Croft 2001:312).

(32) Conceptual Space of Russian Verbs in Terms of RD

P REFERENTIAL DISTANCE

A REFERENTIAL DISTANCE

Data points labeled: aF, pW, pO, aO, aW, oF, oG, pG, oO, oW, aG, pF

	General passives		Actives	a = active, t = typical passive
	Opportunist passives			o = opportunist, p = possessive
	Possessive passives			O = overall, W = write,
	Typical passives			F = forget, G = give

Here, the RD data from the table in (31) has been plotted in two dimensions *à la* Croft, and the axes run from high topicality to low topicality, with high topicality starting in the upper left-hand corner for both axes. This allows for not only more ease in comparing results with Croft, but also for visualizing differences in the numbers of (31) generally. The data points are labeled by verb form for ease of reference. The lowercase letter which marks each data point refers to the voice type: "a" marks each data point as an active, "o" as an opportunist passive, and "p" as a possessive passive. Individual averages for each verb and overall is indicated by the following capital letter. "W" shows the averages of *pisat'/napisat'* 'to write,' "F" the averages of *zabyvat'/zabyt'* 'to forget,' and "G" those for *davat'/dat'* 'to give.' "O" indicates the averages overall for all three verbs. Because of the scale of this particular diagram, I have not labeled the "typical" passives' data points, nor those of the general

average for passives, but I will use a similar scheme in marking these points for passives below (37).

The passive type I am calling "possessive" passives refers to use of participles with which the logical subject is expressed by the preposition *u* + Genitive case.

(33) a. Krome   Gosudarstvennogo, u   neë   napisany
Besides   state   at   her.GEN write.PERF
eshchë   chetyre   gimna.
another   four   hymns
'Besides the National [Anthem], she has written four other anthems.'

b. U   menia   dan   pochti   doslovnyi   eë
At   me.GEN   give.PERF   almost   verbatim   her/its
perevod.
translation
'I have given her/its translation almost verbatim.'

Note that the possessive passives occupy the same general region of our conceptual space as do actives. This is not unexpected, given their definition: to be classified as a "possessive passive," the passive participle would have to be accompanied by an A expressed by the *u* + GEN construction. As most of the Russian passives were of the short variety, lacking an overt A, they fall much lower on the vertical axis. Note as well the very small number of tokens of this type of passive; we should be careful to infer too much from this small a sample.

"Opportunists" is the label I have come up with for those passives whose logical subject is expressed by some modifier to the logical object, *e.g.,* an adjective or genitive modifier.

(34) Lennonovskaia   "Come Together"   byla   napisana
Lennon's.NOM   come together   was   write.PERF

kak   gimn   ego   predvybornoi   kampanii.
like   anthem   his.GEN   election.GEN   campaign.GEN
'Lennon's "Come Together" was written as if to be the
anthem of his election campaign.'

Like the possessive passives, the opportunists' arguments are
more like those of actives in terms of the conceptual space they
occupy. Like the possessive passives, the definition of the
opportunists ensures that they pattern more like actives in that As
must be present for it to be obvious that the passive was encoding the
A in a non-canonical way. Still, while the mere presence of overt As
corresponds to higher placement on the vertical axis, the axes
measure topicality, so it is nevertheless interesting that these types
behave more like actives. The opportunist passive type is also more
well attested than the possessive passive type in the corpus.

Examining (32), we see that the typical passives of all three verbs
are all clustered together at the bottom of the Y axis, while the actives
(a), opportunist (o) and possessive passives (p) are quite distinct. The
placement of the active verbs is patterned like Croft's own placement
of actives with respect to passives in (5) and (8), and is consonant
with his universal (6), repeated here as (35):

(35) If there is a contrast between a basic and non-basic voice. . .
      then the semantic map of the basic voice will include the
      upper right corner of the conceptual space in [(4)], and the
      semantic map of the nonbasic voice will include the lower
      left corner of the conceptual space in [(4)].

                                                    (Croft 2001:315)

The actives in (32) occupy the upper right corner of the conceptual
space, while the passives, the non-basic voice, are lower and to the
left.

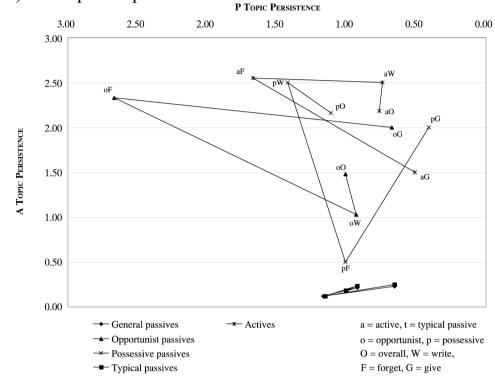One thing we could note, however, is that the passives are not *far*

to the left of the actives. While the difference along the Y axis – itself reflecting the topicality of A, as measured in terms of RD – is significant, the difference along the X axis is much less significant. In fact, in the case of the plot of active forms of *zabyvat'/zabyt'* 'to forget' (marked "aF" on the chart), we are even farther to the left than any of the typical passive forms. Perhaps this has something to do with the meaning of 'forget,' since the thing forgotten is not much affected by the act of forgetting.

There could be several reasons for the seeming discrepancy between Croft's placement of actives and passives, and the location of the Russian actives and passives as plotted in (32). First off, it is not clear what scale should be used for the axes. RD is supposed to be a measure of salience and topicality. A '1' should be maximally topical and a '20' is the ceiling for the value, and should signify maximally untopical. Beyond that, we have no way to be sure that the scale should be perfectly linear; perhaps the difference between a '10' and a '16' is less important than that between a '19' and a '20.'

It could also be that the two axes do not contribute equally to "topicality." According to Croft's space diagrams, passives and actives should appear in opposite corners. In (32), passives and actives appear on opposite ends of the Y axis, but passives generally stand but a little to the left of actives on the X axis. This could signal a problem with Croft's generalizations regarding space (or more likely, their imprecision), or could be indicative of one axis being more important to the overall notion of "salience" than the other.

Finally, it may well be the case that the exact placement of such voice constructions is language-specific, and that for written Russian, passives and actives differ little in the overall topicality of Ps. When we examine data from other languages in section 5.1, I will argue for this approach.

Doing the same plot of relative topicality, but using TP to measure the axes instead of RD, yields the conceptual space in (36):

(36) Conceptual Space of Russian Verbs in Terms of TP

P Topic Persistence

Overall, the results are quite similar. We see that the actives and passives occupy approximately the same areas of the diagram as they did in (32), although the position of actives overall ("aO") and actives for *pisat'/napisat'* 'to write' ("aW") have changed their positions relative to one another. We will return to these differences later.
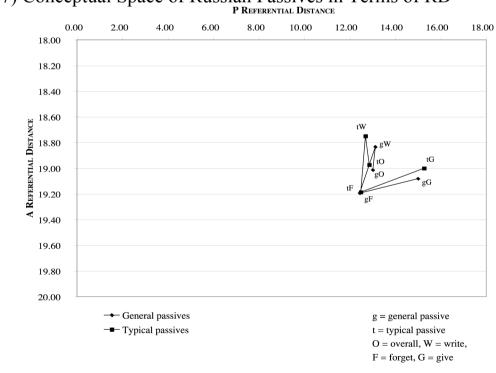
The scale is again arbitrary. The axes in (32) run from 0 to 20, as the measure of RD runs from 1 to 20. In (36), the scale runs from 0 to only 3, and not the 10 which should mark the ceiling for this score. This value was chosen only to afford more granularity in the diagram and thus make the various data points easier to distinguish.

Still, I note with satisfaction that the choice of RD or TP does not generally seem to grossly affect the patterns of the verbs in question. Given that RD is measured in terms of censored data, we might have expected TP to be a more accurate measure of salience than RD and to have provided very different results. The similarity of distribution

suggests that this need not be a concern.

While it is true that we do not know how to compare the axes in terms of RD and TP (e.g., does a '20' RD equal a '0' TP?)—and therefore plotting the verbs in both diagrams does not ensure the verbs will occupy the same general areas of the diagrams—the relative positions seem to be similar regardless of measure. While RD is provided based on censored data and thus statistically somewhat problematic, it is likely that this mathematical truth has little impact on linguistic inquiry. Whether the RD is 20 or 10000, perhaps to human cognition the difference is irrelevant. The referent is simply maximally non-topical or absent.

Before we look at the individual active verb forms, let us look at the typical passives in a little more detail. Because of the scale of (32) and (36), it is hard to distinguish between the different passives plotted on them. Consider (37), which is the same as (32), but with the (vertical) scale changed by omitting the plot of the active data:

(37) Conceptual Space of Russian Passives in Terms of RD

Here we can see the difference between the passive forms a little more clearly, although they are not large differences. Here, the typical passives (t) pattern very much like those of the possessive passive (p) in (32). The verb *pisat'/napisat'* 'to write' occupies an area above and to the left of the overall average ("tO"). The verb *zabyvat'/zabyt'* 'to forget' lies below and to the right of the overall average, while *davat'*/dat' 'to give' lies almost directly to the right of the overall average.

Finally, I present the passive forms by themselves, this time as plotted by A TP and P TP:

(38) Conceptual Space of Russian Passives in Terms of TP

P TOPIC PERSISTENCE

Note that among the three passives examined, all seem to show similar patterns:

The verb *pisat'/napisat'* 'to write' is characterized by lower P salience and higher A salience than the average overall, except when we look at the RD score for actives. Here, the RD for A is a little

higher (and thus the salience a little lower) than the average overall. But looking at TP, we see again that A is a bit higher in salience than the average overall.

Despite variance in A salience as measured by RD versus TP, the verb *davat'/dat'* 'to give' is marked by higher P salience than is *pisat'/napisat'*, its average being to the right in the conceptual space.

Whether measured by RD or TP, the verb *zabyvat'/zabyt'* 'to forget' consistently has higher A topicality and lower P topicality than either *pisat'/napisat'* or *davat'/dat'*, and is placed lower and to the left of these verbs in the conceptual space.

Moreover, these patterns are very similar for the active counterparts of these passives. The active forms of *pisat'/napisat'* 'to write' inhabit a region similar to the passive forms, relative to *davat'/dat'* 'to give.' While the active forms of *zabyvat'/zabyt'* 'to forget' show a higher A topicality than do the passive forms of *pisat'/napisat'* 'to write' and *davat'/dat'* 'to give,' active forms of *zabyvat'/zabyt'* 'to forget' are still marked by a lower P topicality relative to these two verbs.
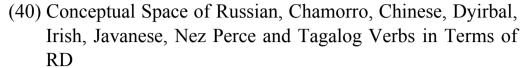
It would seem that the profile for the arguments of any of these three verbs tends to place them in areas close to, but distinct from, the other verbs examined. This placement seems to be consistent, regardless of the actual form or voice of the verb. It is tempting to attribute these facts to differences in the semantics of the verbs in question.
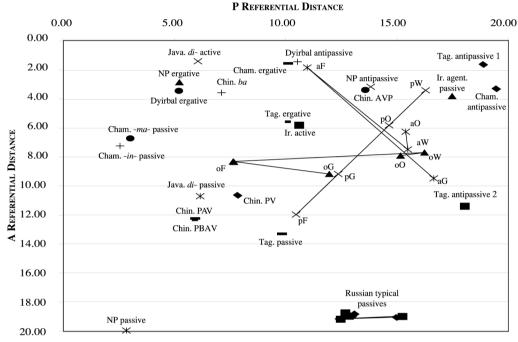
5.1. Language-specific Constructions in Conceptual Space

Earlier I noted that it was not immediately clear how to compare constructions across languages based on RD and TP. Nevertheless, Croft's assertion is that plotting language-specific constructions in such conceptual space is the only valid way to compare them. To that end, I here attempt to use data from studies of topicality in other languages:

(39) Quantitative Studies of RD and TP Data Used in Forming
    Conceptual Space

Language	*Verb forms/voice*	*Source*
Chamorro	active	Cooreman (1984)
	-*ma*-passive	
	-*in*-passive	
Chinese	agent-verb-patient (AVP) construction	Xing (1993)
	bǎ construction	
	patient *bèi* agent verb (PBAV)	
	patient agent verb (PAV) construction	
	patient verb (PV) construction	
Dyirbal	ergative	Cooreman (1988)
	antipassive	
Irish	active	Noonan (1990)
	passive	
Javanese	'active' *di*-construction	Myhill (1992)
	'passive' *di*-construction	
Nez Perce	ergative	Rude (1988)
	antipassive	
	passive	
Tagalog	ergative	Cooreman et al. (1984)
	passive	
	antipassive (i)	
	antipassive (ii)	

First, here is our conceptual space again as measured in terms of
RD, adding the data points to represent the data from (39):

(40) Conceptual Space of Russian, Chamorro, Chinese, Dyirbal, Irish, Javanese, Nez Perce and Tagalog Verbs in Terms of RD

We see that the language-specific constructions are scattered all over our conceptual space, their ranges apparently overlapping.

This fact is in keeping with a prototype analysis of such voice constructions. We expect fuzzy boundaries to categorization rather than hard-and-fast rules dictating membership within a group. Also, recall that what we are plotting here is essentially a probabilistic tendency: a likelihood that a given topicality profile for an A or P will result in the use of one construction over another.

The prototype analysis also accounts for why Myhill characterizes the various Chinese patient-verb constructions (the patient *bèi* agent verb (PBAV) construction, patient agent verb (PAV) construction, and patient verb (PV) construction) as serving "a discourse function somewhat like that of a passive construction" and that "the *bǎ* -structure appears to serve a discourse function somewhat like that of an ergative or inverse construction" (Myhill 1992:160-1).

Note the placement of the patient-verb constructions in the same general region as the Javanese and Tagalog passives, and the *bǎ* construction between the Tagalog, Dyirbal and Chamorro ergatives. Their proximity to these other constructions should signal similar discourse function. The PBAV, PAV, and PV constructions also all stand in relation to the AVP and *bǎ* constructions as do passives to actives: lower and to the left.

The distribution also supports Croft's assertion that language-specific constructions would occupy different points of his conceptual space; we note that while many of the passive constructions, for example, tend to congregate toward the lower left-hand corner of the conceptual space, they are scattered among different areas.

The actives for Russian, Javanese and Irish all occupy similar areas of the conceptual space, particularly the Javanese and Irish actives and the active forms of Russian *zabyvat'/zabyt'*. What is interesting is the different placement of the passives from these three languages.

We find that the Chamorro passives do not occupy the space that Croft portrays in diagrams like (8), but they inhabit regions higher along the vertical axis. In fact, the Chamorro passives are roughly parallel with the Russian active forms along the vertical axis. The Chamorro passives are also higher than the typical Russian passives along the vertical axis, and farther to the left on the horizontal axis. That is, the Chamorro passives are marked by lower A RD and P RD than are the Russian passives. Still, compared to the Chamorro ergative, these passives are, per Croft's description, lower and to the left in our conceptual space.

The relationships between the Javanese active and passive, and the Tagalog ergative and passive, are similar to that of the Russian active and passive; there is a marked difference in A RD, but little variation at all in P RD. The difference is almost entirely vertical. The Nez Perce ergative and passive differ a little more horizontally, but otherwise much the same can be said of the relationship between the

Nez Perce constructions.[12]

Noonan's written Irish data is most striking, in that the Irish passive behaves much more like an active form that it does a passive. It is marked by lower A RD and higher P RD than are the Irish actives. It is not surprising that Myhill (1992) points to the Irish passive as an "exceptional case." While this construction may have some structural characteristics that suggests a characterization as "passive" – namely, the P of a corresponding sentence expressed as a subject, while the A of a corresponding sentence is expressed as an oblique – it is clear that functionally, this construction does not behave like a passive. Further evidence of its exceptional discourse function is provided by the fact that active forms barely outnumber the agentive passives in Noonan's counts (Myhill 1992:111-2).
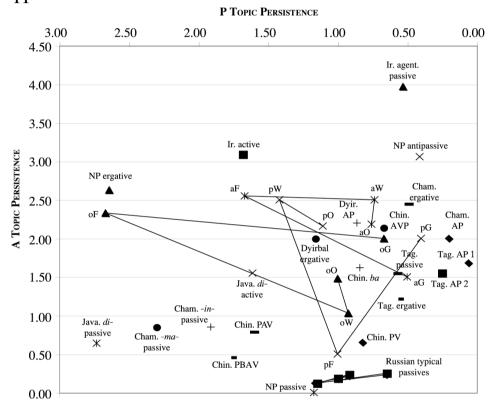
The antipassives for Nez Perce, Dyirbal, Tagalog, and Chamorro all seem to occupy an area to the right and above the active constructions, again roughly conforming to Croft's characterization in (8) of the antipassive occupying conceptual space to the right and above the active/direct construction. The sole exception seems to be the Tagalog "antipassive 2" construction, which is to the right of the active/direct, but much lower than the Tagalog "antipassive 1" construction. The A RD for this construction is much lower, being lower than that of the Tagalog ergative and similar to that of the Tagalog passive.

Ergatives, which are not addressed by Croft in (8), seem to occupy an area slightly above and to the left of the actives, though we will see that this relationship does not seem to hold when looking at topicality in terms of TP. At any rate, the relationship of each language's ergative to the other voice forms within the language are probably more informative than comparison of ergatives to active/direct constructions.

---

12 Note that the Nez Perce passive is always short, disallowing A altogether, so the RD of A is always 20.

Looking again at the topicality of arguments for the languages in (39) in terms of TP, we arrive at (41):

(41) Conceptual Space of Russian, Chamorro, Chinese, Dyirbal, Irish, Javanese, Nez Perce and Tagalog Verbs in Terms of TP

While the Russian data more or less "held its place" in the transition from RD and TP—the active and passive forms for Russian seem to occupy the same general areas whether conceptual space is measured in terms of RD or TP—we see more variation for some of the other language-specific constructions.

The actives occupy an area which might be characterized as "up and to the right," but their positions relative to each other have changed. The Javanese active is lower along the vertical axis in

relation to say, the Russian active overall, or the Chinese AVP construction. The Irish active, on the other hand, is higher along the same axis than are the Russian active or Chinese AVP construction, when it was lower along the vertical axis in terms of RD.

The Chamorro passives still inhabit a region to the left of Russian actives and passives, and above that of the Russian passives. This suggests that P salience is more important for Chamorro passives, and A salience is more important for Russian passives. Again, compared to the Chamorro ergative, these passive constructions are lower and to the left in our conceptual space.

Likewise, the Javanese active and passive constructions maintain their relative positions; the Javanese passive is lower and to the left of the Javanese active. Seemingly anomalous is the relationship between the Tagalog ergative and Tagalog passive, as the ergative construction is now to the left and below the passive. The relationship between the Nez Perce ergative and passive is similarly convoluted, in that while the passive has much lower A TP than does the ergative, it also has lower P TP.

The Irish data is interesting in that both the active and passive are marked by considerably greater A TP than any active or passive of the other two languages. This puts them well above the active conceptual space for Russian actives, unlike the RD measures of (40). Still, the two Irish constructions maintain their overall relationship to each other, and the Irish passive is still notably in the upper right-hand corner of the diagram, contrary to the characterization of passive.

Under the TP measure, the antipassives linger in the upper right-hand corner of the diagram, and the Chinese PBAV, PAV, and PV constructions still inhabit the same regions as passives. Their relationship to other voice constructions within the same language are generally unchanged.

While I think it is possible to see that the relative relationship between these constructions is similar regardless of which topicality measure is chosen (RD or TP), it is nevertheless clear that the plots

for some of these constructions is different. Both of these measures are supposed to be gauges of topicality for arguments, but in precisely what way we cannot say. To attempt to consolidate our characterizations of conceptual space, let us introduce a final conceptual space, based on a combinatory measure which I will call the *Salience Rating* (42):
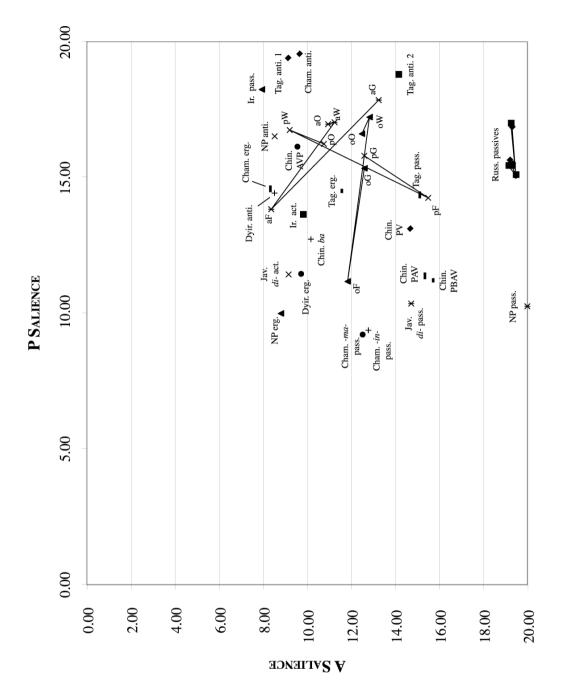
(42) Salience Rating (SR)

$$SR(x) = \frac{RD(x) + (20 - 2 \times TP(x))}{2}$$

where RD = Referential Distance of the argument,
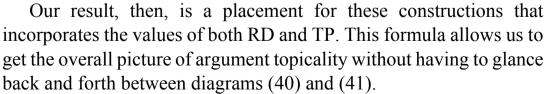   TP = Topic Persistence of the argument

The Salience Rating (hereafter, SR) is simply a mechanism by which I will "integrate" RD and TP.[13] Whereas RD runs from 1 to 20 in descending topicality, TP runs from 0 to 10 in ascending topicality. So in (42), TP is converted to run from 0 to 20 by multiplying it by 2, and then the result is subtracted from 20 to "reverse" the topicality. The two numbers are then averaged, resulting in a number ranging from 0.5 to 20. Like RD, the higher the SR, the lower the overall topicality of the argument.

Finally, we plot the conceptual space of the language-specific constructions in (39) in terms of this new measure, (43):

(43) Conceptual Space of Russian, Chamorro, Chinese, Dyirbal, Irish, Javanese, Nez Perce and Tagalog Verbs in Terms of SR

---

[13] This formula gives equal weight to RD and TP. It could just as easily give more weight to one or the other measures, if there were ample reason to do so. For example, if a later study were to show that the censored data of RD was less accurate than that of TP, TP could be given precedence.

Our result, then, is a placement for these constructions that incorporates the values of both RD and TP. This formula allows us to get the overall picture of argument topicality without having to glance back and forth between diagrams (40) and (41).

I believe these results allow us to broadly maintain Croft's

generalizations. The general position of the passives in all of these languages is lower and to the left of the active constructions, with the notable exception of the problematic Irish passive. The ergatives cluster in approximately the same place as the actives, but slightly to the left, suggesting higher P topicality. The antipassives occupy a space above and to the right of the active forms.

6. Conclusion

Constructions that are labeled as "passives" appear in many languages; what is problematic is that linguists cannot agree on what it *means* to be a passive. There are obviously some similarities between the English passive, the Turkish passive and the Irish passive which lead to these constructions being labeled as "passives." But most similarities we might note fail to apply to *all* constructions labeled as passive in the literature.

Siewierska (1984) notes that there are no structural features shared by all passives, cross-linguistically, and Croft (2001) has suggested that the phenomena of all voice constructions are actually part of a continuum of constructions. Just as a dialect continuum makes drawing a line between one language and another an arbitrary decision, deciding what is or is not a passive will depend on what one's definition of passive is.

This is an advantage of the prototype analysis of the passive, common to cognitive linguistics (such as Langacker (1982), Shibatani (1988), Arnett (1995), Goldberg (1995)). Membership within a prototype category is gradient and probabilistic, and therefore "fuzzy." Categorization is a matter of possessing features which weight towards a membership threshold, but which features are salient may vary from member to member. For passives, we can note features which are more likely to be possessed by passive constructions, but we also recognize that these features may not be

shared by all members of the category "passive." Those constructions which are "better examples" of the passive are simply those with more central or salient features of the prototype.

6.1. Conceptual Space of Constructions

I introduced RD and TP data for As and Ps of three Russian verbs (*pisat'*/*napisat'* 'to write,' *zabyvat'*/*zabyt'* 'to forget,' and davat'/*dat'* 'to give') in section 4. Croft's (2001) notion of comparing language-specific constructions in a conceptual space defined by A and P topicality was done somewhat subjectively: either as a continuum running from "subject-like" to "object-like," or from "salient" to "absent." I substituted the measures of RD, and TP (and the integration of the two, which I termed the "salience rating" (SR)) as a means of more objectively approximating the abstract notion of "topicality." The results showed that each of the Russian verbs occupied similar, but distinct, areas of the conceptual space. Their relative positions were similar between active and passive forms, and seemed consistent whether the measure of RD or TP was used.

Comparing the position of the Russian verbs in this space to constructions in other languages, the constructions demonstrated similar relations to those suggested by Croft, and largely seemed to be internally consistent; that is, while passives from the various languages examined did not all inhabit the exact same regions of the conceptual space, they did tend towards the lower left-hand corner of the space, and generally stood lower and to the left of any active or ergative constructions within the language in question.

One of Croft's arguments against the universal characterization of passive was that language-specific passives inhabit different regions of the conceptual space mapping A and P topicality. The disparate results I presented in section 4 for the topicality of various voice constructions in Russian, Chamorro, Chinese, Dyirbal, Irish, Javanese, Nez Perce and Tagalog are in keeping with Croft's belief

that the way to compare voice constructions cross-linguistically is through such diagrams.

Moreover, the Russian data suggest that the continuum of constructions that can be plotted in such space cross-linguistically also extends to individual verbs within a given language. The topicality of As and Ps for the three Russian verbs studied within this work is consistent and places each verb within its own area of our conceptual space.

The overall model is one of a conceptual space that is divided between different voice constructions. While speakers are free to choose whatever construction they want, it is assumed that they will be motivated to make grammatical utterances in their effort to communicate. Likewise, it is assumed that speakers will utilize felicitous utterances, in following the pragmatic considerations of their language. Choice of voice constructions like passive will be motivated not only by the semantics of the verb in question, but by the topicality of the arguments of the verb.

Essentially, Croft proposes that while speech communities will form their own preferences as to how topicality affects the choice of voice, there are universal principles guiding the division of this space among voice constructions, such as (6). Beyond these generalizations, languages are free to parcel up the conceptual space in any of a variety of ways. This fact makes room for variation among individual languages and individual constructions within each language and different placement within our conceptual space.

Along the same lines, speech communities will "decide" which factors serve to distinguish voice constructions in their language. For Russian, it appears as if the topicality of the A is most important in the choice between active and passive voice. This also appears to be the case in Nez Perce and Javanese. In Chamorro, it would appear that P topicality plays a larger role in the selection of active or passive voice.

The overlapping nature of the placement of these constructions in

our conceptual space and the probabilistic nature of voice selection are amenable to a prototype analysis. We expect the boundaries between the various constructions to be flexible, and the presence of certain features to be more indicative of one voice or another, but not prescriptive. The same situation has probably led to some of the confusion among linguists as to how to classify constructions, or even as to what a universal characterization of passive should be. There are likely a combination of structural and functional characteristics exhibited by languages to distinguish between voices; linguists have seized upon any number of these as defining the passive. In reality, languages will parcel the conceptual space using some of these features, but not others, and each language may end up with a different set of features. The result is a continuum of constructions, whose membership within a category can really only be understood in terms of a prototype.

References

Allen, R. 1978. *The Perfect in Russian Dialects*. Master Thesis, Chapel Hill, NC: University of North Carolina.

Arnett, C. 1995. *The Passive in German: The View from Cognitive Grammar*. Doctoral Dissertation, Ann Arbor, MI: University of Michigan.

Clancy, S. 2000. *The Chain of BEING and HAVING in Slavic*. Ph.D. Dissertation, Chapel Hill, NC: University of North Carolina.

Cooreman, A. 1984. A Functional Analysis of Passives in Chamorro Narrative Discourse. *Papers in Linguistics* 16, 395-428.

_____. 1987. *Transitivity and Discourse Continuity in Chamorro Narratives*. Berlin: Mouton de Gruyter.

_____. 1988. Ergativity in Dyribal Discourse. *Linguistics* 26, 717-46.

Cooreman, A., B. Fox, & T. Givón. 1984. Ergativity in Dyribal Discourse.

Studies in Language 6, 343-74.

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

Gattnar, A. 2002. Russian Corpora in Tuebingen. University of Tübingen. 12 September 2002 <http://www.sfb441.uni-tuebingen.de/b1/en/korpora.html>.

Givón, T. 1983. *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam: Benjamins.

Goldberg, A. 1995. *Constructions: A Construction Grammar Approach*. Chicago, IL: University of Chicago Press.

Janda, L. 2003. Linguists for the Responsible Use of Internet Data. Chapel Hill, NC: University of North Carolina. 28 October 2003 <http://www.unc.edu/ lajanda/responsible.html>.

Keenan, E. 1985. Passive in the World's Languages. In T. Shopen (ed.), *Language Typology and Syntactic Description, Volume 1: Clause Structure* 243-281. Cambridge: Cambridge University Press.

Langacker, R. 1982. Space Grammar, Analysability, and the English Passive. *Language* 58.1, 22-80.

Launer, M. 1974. *Elementary Russian Syntax*. Cambridge, MA: Slavica.

Lönngren, L. 1993. *Chastotnyi Slovar' Sovremennogo Russkogo Iazyka*. Uppsala: Acta Universitatis Upsaliensis.

Myhill, J. 1992. *Typological Discourse Analysis: Quantitative Approaches to the Study of Linguistic Function*. Oxford: Blackwell.

Noonan, M. 1990. Tale of Two Passives in Irish. Paper Presented at the Symposium on Voice, Santa Barbara, CA.

Rude, N. 1988. Ergative, Passive, and Antipassive in Nez Perce: A Discourse Perspective. In M. Shibatani (1988), 547-60.

Shibatani, M. (ed.). 1988. *Passive and Voice*. Amsterdam: Benjamins.

Siewierska, A. 1984. *The Passive: A Comparative Linguistic Analysis*. London: Croom Helm.

_____. 1993. Syntactic Weight vs. Information Structure and Word Order Variation in Polish. *Journal of Linguistics* 29, 233-265.

Thode, H. 2002. *Testing for Normality*. New York: Marcel Dekker.

Thompson, C. 1994. *Passive and Inverse Constructions*. Amsterdam: Benjamins. 47-63.

Xing, Z. 1993. Discourse Functions of Word Order in Chinese: A Quantitative Analysis of Diachronic Texts. Doctoral Dissertation, Ann Arbor, MI: University of Michigan.