Transitivity in RSL: a Corpus-Based Account

Vadim Kimmelman
ACLC, University of Amsterdam
Amsterdam, the Netherlands
v.kimmelman@uva.nl

Abstract

A recent typological study of transitivity by Haspelmath (2015) demonstrated that verbs can be ranked according to transitivity prominence, that is, according to how likely they are to be transitive cross-linguistically. This ranking can be argued to be cognitively rooted (based on the properties of the events and their participants) or frequency-related (based on the frequency of different types of events in the real world). Both types of explanation imply that the transitivity ranking should apply across modalities. To test it, we analysed transitivity of frequent verbs in the corpus of Russian Sign Language by calculating the proportion of overt direct and indirect objects and clausal complements. We found that transitivity as expressed by the proportion of overt direct objects is highly positively correlated with the transitive prominence determined cross-linguistically. We thus confirmed the modality-independent nature of transitivity ranking.

Keywords: argument structure, transitivity, Russian Sign Language

1. Introduction

1.1. Theoretical Background

Transitivity is the possibility of a verb to occur with a direct object.[1] If we look at transitivity cross-linguistically, we can observe that different verbs are transitive to a different degree. In a recent study, Haspelmath (2015) analysed a sample of 36 languages in order to assess transitivity cross-linguistically. For each language, the possibility of using a transitive frame has been collected for a list of 70 verbs. Two main results came out of this study: (1) languages differ with respect to the proportion of transitive verbs they have and (2) verbs differ with respect to the proportion of languages in which they are transitive. This latter result generates a ranking of verbs by transitivity prominence, such that some verbs like *die* or *jump* are never transitive, other verbs like *break* and *tear* are transitive in all languages in the sample, and some verbs like *leave* or *know* are transitive in some languages but not others.

Several explanations have been offered for the fact that some verbs are more likely to be transitive than others. For instance, Malchukov (2005) proposed a semantic map for transitivity as in Figure 1. The verbal meanings to the left are more likely to be transitive than the verbs to the right. Verbs of effective action (such as *break*) are the most likely to be transitive because they involve an Agent (an active and volitional participant), who has an effect on the Patient (and the Patient changes as a result of the action). Verbs of contact also involve an Agent and a Patient, but the Patient is not changed by the actions of the Agent, and with verbs of pursuit there is also no contact between the Agent and the Patient. Verbs of perception, emotion, and sensation constitute a separate sub-part of the map because there the notions of Agent and Patient are less directly applicable, and one might speak of an Experiencer and a Stimulus, so they are less likely to be transitive than verbs of effective action, but not ordered with respect to verbs of contact or pursuit. Such a cognitive explanation of transitivity is likely to apply across modalities, so we expect to find similar patterns of transitivity in sign languages, too.

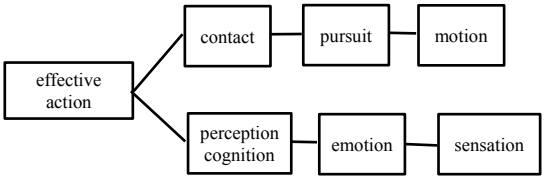

Figure 1: Malchukov's semantic map for transitivity.

Another explanation for the differences in transitivity between different verbs is frequency-based (Haspelmath, 1993; Haspelmath et al., 2014). It is intuitively true (albeit probably not explicitly tested) that some actions in real world frequently involve two participants (such as breaking), while others frequently involve one participant (such as freezing), so the former are more likely to be lexicalised as transitive and the latter as intransitive. Again, such an explanation for transitivity is likely to apply across modalities, so also to sign languages.

The two explanations are not incompatible, as for instance verbs of effective action describe the situations which necessarily frequently involve two participants. However, the frequency-based explanation makes a clear prediction with respect to transitivity in Russian Sign Language (RSL).

Although transitivity ranking is defined cross-linguistically, we think that there is a way of applying transitivity ranking to one language and testing it based on corpus data. We predict for an individual language (signed or spoken) that the frequency of overt object expression should reflect the transitivity prominence from Haspelmath (2015): If a particular event is more likely to have two participants in real

[1] By this definition we consider ditransitive verbs to be a subclass of transitive verbs: they have a direct object, and also an indirect object

world, the verb used to describe this event is more likely to be lexically transitive cross-linguistically AND to have an overt object in corpus data of a particular language.

At first sight it might seem that overt objects in corpus data do not directly reflect transitivity of the verbs, because even transitive verbs can occur without overt objects in languages which allow object pro-drop (and RSL is such a language). However, the possibility of pro-drop is orthogonal to the notion of transitivity. Imagine a language that has transitive verbs X and Y which are transitive to a different degree, so that X is sometimes used intransitively[2], and Y is always (semantically) transitive[3]; object pro-drop in this language is allowed. We then expect that pro-drop will cause X and Y to appear without objects in a certain percentage of cases (a%), but we expect the effect of pro-drop to be the same for the two verbs, because pro-drop depends on the definiteness of the object, and not on its presence in the lexical argument structure of the verb. However, the proportion of cases where the verb is indeed used intransitively and no object is implied depends on the verbal meaning of X and Y and reflects real world frequencies of the events with one or two participants, so it would be different for X and Y (say x% and y% of all cases). The observed percentages of overt objects for X and Y will be (100-(a+x))% and (100-(a+y))% respectively, and so despite the presence of cases of pro-drop, these percentages will also reflect transitivity of these verbs.

1.2. The Current Study

In order to test the relation between the transitivity prominence and overt object expression, we analysed corpus data of RSL. RSL is a language used by at least 120 000 people in Russian Federation. Although some research on linguistic properties of RSL has recently appeared, no studies of argument structure are available yet. Importantly, corpus-based investigations of transitivity have not yet been conducted for any other sign languages.

Apart from providing a basic corpus-based description of transitivity in RSL, in this paper we are interested in answering two more general questions: (1) Does transitivity prominence apply across modalities? and (2) Is transitivity prominence related to overt object expression?

2. Methodology

The RSL corpus (Burkova, 2015) contains naturalistic narratives and a small number of dialogues produced by 37 signers from different regions of Russia. The total length of the video files is 5h 28min. The whole corpus is annotated in ELAN with sign-by-sign translation (separately for the right and left hands) and with sentence-level free translation in Russian; the total number of annotations is over 65 000. Based on the number of annotations on the right hand tier, the estimated number of signs in the corpus is ≈25 000.

To study transitivity, verbal tokens expressing the 80 typical verbal meanings from the VALPAL project (Hartmann et al., 2013) have been identified.[4] Since the RSL corpus does not make use of lexical IDs, the same sign is sometimes translated with different glosses, and the same gloss is sometimes used for different signs; therefore, the search for tokens involved identifying all possible patterns with which a particular verbal meaning would be expressed in Russian, and using regular expressions to search for them. For instance, the meaning 'to dress' in Russian can be expressed by two verbs with different prefixes: *o-devat'* and *na-devat'*, and also the gloss might contain the corresponding noun *odezhda*, so a regular expression was used to identify all possible tokens.

Sometimes a meaning from the VALPAL list was matched to more than one sign in RSL, so labels like SPEAK1, SPEAK2, SPEAK3 have been created to reflect that different signs have the meaning 'to speak'. At first, all different forms were assigned different labels. However, sometimes the differences between the forms appear to be phonetic. For instance, RUN1 contains an alternating movement of the hands, RUN2 contains an identical but synchronized movement of the hands, and RUN3 contains the same movement as RUN2 but only involving fingers. In such cases we also looked at the argument structure of the different variants (as described below) in order to decide whether they should be analysed as one verb. In most cases it turned out that different variants also differed in argument structure, as was the case with SPEAK1, SPEAK2, and SPEAK3, but for the verbs RUN and COME the variants were not different, thus for further analysis we grouped them together.

We excluded the meanings that are expressed in RSL by classifier predicates, as classifier predicates have modality-specific properties, crucially also with respect to argument structure (Benedicto and Brentari, 2004; Kimmelman et al., 2016), and have to be analysed separately. However, we included two signs which can potentially be classifier predicates, namely GIVE and TAKE. In our data there are examples in which these signs are used with abstract objects (1) or with objects that do not correspond to the handshapes used in the predicates, so it is clear that these verbs at least have non-classifier usages. Note, however, that we then included all usages of these verbs, so some tokens of these verbs might in fact be classifier predicates.

(1) MORE INFORMATION GIVE
'[...] to give more information.'

Altogether, based on the 80 verbal meanings, we annotated 117 verbal signs (types) totalling 2248 tokens. For each token we created annotations on several tiers, including a tier for the verbal meaning and a tier for word order in the clause. Clause boundaries were identified on semantic grounds: the verb and all its arguments and adjuncts (similar to Hansen and Hessmann (2008)). We labelled the most agentive argument as S. For verbs which only take one object, we labelled the relevant argument as O (direct object). For ditransitives, such as SAY, SPEAK, TELL, TALK, TEACH we labelled the Addressee/Goal as O and the

[2] As for instance transitive verbs *to melt* and *to eat* can be used intransitively in English.

[3] As for instance the transitive verb *to give*, which can only be used transitively in English.

[4] The VALPAL project used a list of 80 verbal meanings; however, Haspelmath's transitivity prominence is only discussed for 70 of the 80 verbs in Haspelmath (2015).

Theme/Patient as O2. This is an unusual decision from a typological perspective, as the Theme/Patient is usually the direct object. However, in RSL all these verbs are agreeing and they agree with the Addressee/Goal argument, so based on the language-internal criterion of agreement we identified this argument as the direct object. We also identified possible clausal complements CO in cases like *He saw [that she left]*.

In order to compare transitivity of RSL verb to Haspelmath's transitivity prominence ranking, we calculated proportions of overt arguments per verbal type (e.g. number of tokens with overt O divided by the total number of tokens). For the sake of completeness, we calculated proportions of overt direct and indirect objects, clausal complements, and subjects. Since we are working with proportions, verbal types with a small amount of tokens could distort the picture, so we only included verbs with at least 25 tokens. This resulted in 29 verbal types totalling 1611 tokens. Note that these types includes some groups of verbs which correspond to one meaning in Haspelmath's ranking, such as SPEAK2 and SPEAK3, and also some verbs from the longer VALPAL list of meanings which are not included in Haspelmath's ranking. Thus for the comparison with this ranking we were left with 25 verbal types, but for the general overview of overt expression of arguments we report on all 29 frequent verbal meanings.

3. Results

Basic properties of argument structure of frequent verbs in RSL can be assessed by looking at the overt direct and indirect objects as well as clausal complements which accompany these verbs. In Figure 2, the frequent RSL verbs are ranked according to the proportion of overt direct objects. One can immediately see that this ranking looks quite reasonable, with verbs like HAPPY and LAUGH never occurring with a direct object, and verbs like TAKE and GIVE being the most frequently accompanied with an overt direct object.

Figure 2: Frequent RSL verbs ordered by proportion of overt direct objects.

In Figure 3, the frequent RSL verbs are ranked according to the proportion of overt indirect objects. Again, this ranking intuitively makes sense, as the meanings which are typically ditransitive cross-linguistically, such as TEACH and TELL are high in this ranking.

Finally, in Figure 4, the frequent RSL verbs are ranked according to the proportion of overt clausal complements. Not surprisingly, verbs of speech, perception and cognition are high in this ranking, as semantically the complements of these verbs are often full propositions and not entities.

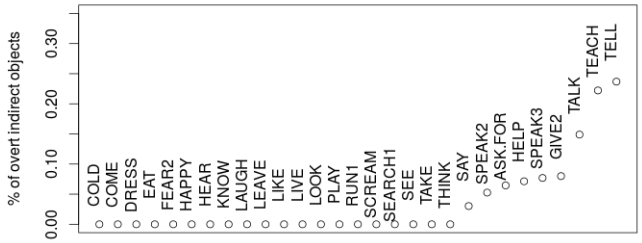

Figure 3: Frequent RSL verbs ordered by proportion of overt indirect objects.

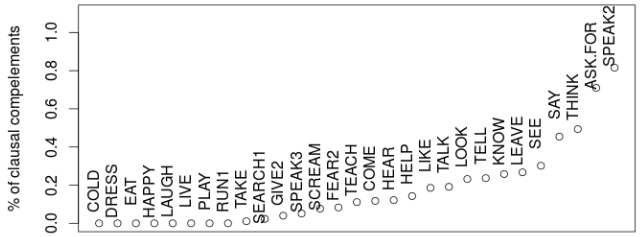

Figure 4: Frequent RSL verbs ordered by proportion of overt clausal complements.

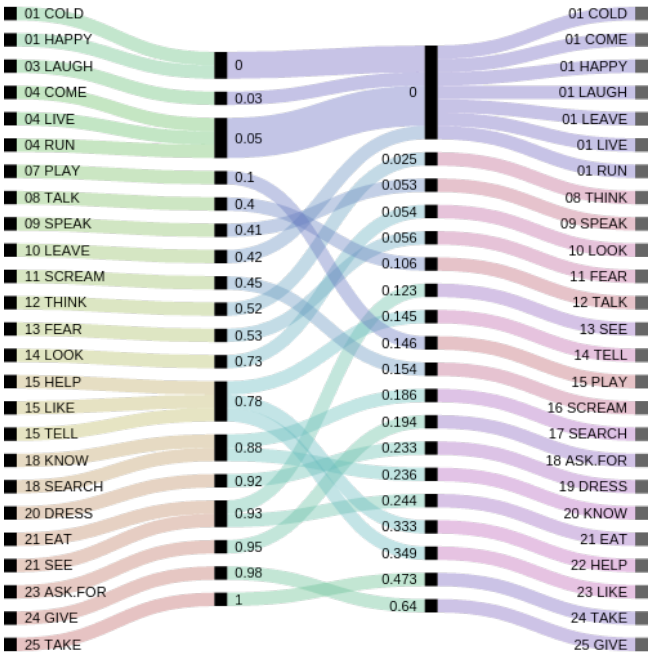

Figure 5: Haspelmath's ranking (left) and ranking based on overt objects in RSL corpus (right). Created with RAW (raw.densitydesign.org).

It should be clear that the best general measure of transitivity is the proportion of overt direct objects, because indirect objects and clausal complements target specific semantic subclasses of verbs. Thus we compared the ranking from Figure 2 to the ranking by transitivity prominence from Haspelmath (2015). For each verb, Haspelmath provided a number which represents the proportion of languages in which a particular verb is lexically transitive. In order to compare our ranking, we calculated the correlation between this measure and the proportion of overt direct objects in

our RSL corpus. Since neither data sets are normally distributed, we used Spearman's rank correlation. It turned out that the two measures are highly positively correlated ($\rho = 0.849$) and that the correlation is highly significant ($p = 8.081 * 10^{-8}$).[5]

The relation between the ranking based on the RSL corpus and Haspelmath's ranking is provided in Figure 5 as an alluvial diagram. On the left, the verbs are ranked according to transitivity prominence (Haspelmath, 2015), and connected to the numerical values of transitivity prominence (center left). On the right, the verbs are ranked according to the proportion of overt direct objects in the RSL data, and connected to the numerical values of the proportion (center right). The numerical values of the corresponding verbs are connected to each other. This figure also clearly represents the ties, so it lets one see for instance the verbs which are never transitive cross-linguistically (COLD, HAPPY) and the verbs which never occur with overt direct objects in RSL (COLD, HAPPY, LAUGH, COME, LIVE, RUN).

From the graph and the high correlation coefficient it should be clear that transitivity prominence as defined by Haspelmath based on cross-linguistic data is also reflected by the proportion of overt direct objects in the corpus data of RSL. For the sake of completeness, we also calculated correlations between the proportions of overt indirect objects, clausal complements, and subjects, and only found low correlations which were not statistically significant. This is in agreement with our expectation, as indirect objects, clausal complements, and (especially) subjects should not reflect transitivity in general.

4. Conclusions

In this paper, we provided a first description of transitivity in RSL (that is, overt expression of objects of different types) based on the corpus of RSL. We have demonstrated that the presence and frequency of objects, indirect objects is related to the verb's semantics, as can be seen in figures 2, 3, and 4 above. RSL is thus compatible with cognitively-based accounts of transitivity, such as in Malchukov (2005). For instance, verbs of effective action such as TAKE are more likely to co-occur with an overt object than verbs of pursuit such as SEARCH or cognition, such as KNOW.

We have also shown that the transitivity ranking based on the proportion of overt direct objects in the corpus correlates well with the transitivity prominence ranking (Haspelmath, 2015). This has two theoretical consequences:

1. Transitivity is a modality-independent phenomenon, as the same verbs are likely to be transitive in spoken and signed languages.

2. The proportion of overt objects in a corpus of a single language reflects transitivity prominence.

The second consequence is also of practical importance. When analysing argument structure in sign languages based on corpus data, researchers are often limited by the lack of grammatical annotations in the existing corpora. However, if a simple measure – the proportion of overt objects – reflects transitivity in general, investigation of the basic properties of argument structure in sign language becomes relatively simple. The only annotations necessary for such an analysis are clause boundaries and labels for predicates and arguments.

In addition, if one finds a verb which, based on the proportion of overt objects, occurs in an unusual place in the ranking judging by comparison between the corpus-based and the transitivity prominence rankings, one might want to further investigate this verb as it might be a typologically exceptional item. For instance, in the RSL data analysed here, the verb SEE is unexpectedly infrequently used with an overt direct object. This can be explained by one of the two factors (or their cumulative effect): (1) this verb often occurs with clausal complements, which do not count as direct objects in our analysis and (2) the meaning of this verb might be closer to 'look' than to 'see' as it might not imply that an image of the object has been attained. Whether the latter explanation applies should be tested in future research.

[5] Due to the presence of ties, the p-value is an approximation.

5. Acknowledgements

I would like to thank Roland Pfau and Marloes Oomen for their comments on this paper. This project has been supported by the NWO (The Netherlands Organisation for Scientific Research), project number 360-70-520.

6. Bibliographical References

Benedicto, E. and Brentari, D. (2004). Where did all the arguments go?: Argument-changing properties of classifiers in ASL. *Natural Language & Linguistic Theory*, 22(4):743–810.

Burkova, S. (2015). Russian sign language corpus. http://rsl.nstu.ru/.

Hansen, M. and Hessmann, J. (2008). Matching propositional content and formal markers: Sentence boundaries in a DGS text. *Sign Language & Linguistics*, 10(2):145–175.

Hartmann, I., Haspelmath, M., and Taylor, B. (2013). *Valency Patterns Leipzig*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Haspelmath, M., Calude, A., Spagnol, M., Narrog, H., and Bamyaci, E. (2014). Coding causal–noncausal verb alternations: A form–frequency correspondence explanation. *Journal of Linguistics*, 50(03):587–625.

Haspelmath, M. (1993). More on the typology of inchoative/causative alternations. In Bernard Comrie et al., editors, *Causatives and transitivity*, pages 87–120. John Benjamins, Amsterdam.

Haspelmath, M. (2015). Transitivity prominence. In Andrej Malchukov et al., editors, *Valency Classes in the World's Languages*, pages 131–147. de Gruyter Mouton, Berlin.

Kimmelman, V., Pfau, R., and Aboh, E. O. (2016). Argument structure of Russian Sign Language classifiers: insights from corpus data. Poster presented at TISLR 12.

Malchukov, A. (2005). Case pattern splits, verb types and construction competition. In Mengistu Amberber et al., editors, *Competition and variation in natural languages: The case for case*, pages 73–117. Elsevier.