Architectural constraints vs. efficiency constraints: Argument coding and verbal voice coding

MARTIN HASPELMATH
MPI-EVA Leipzig

Abstract: In this talk, I discuss several universals of argument coding and verbal voice coding and argue that they can be explained on the basis of efficiency of coding, i.e. a tradeoff between speaker and hearer needs (a functional-adaptive explanation). These constructions have often been discussed in a generative context, where explanations have been expected to come from innate architectural and substantive biocognitive constraints ("UG"). I observe that the generative proposals have not converged, while the coding asymmetries that are the focus of my work have proved robust and ubiquitous in grammar.

1. Theoretical goal: Explaining language universals

My resarch has the ultimate goal of explaining grammatical universals, by identifying general causal factors (***pressures***, or ***forces***, or ***constraints***) that lead to these outcomes.

This is why several of my papers start with "Explaining..."
(e.g. Haspelmath 2004; 2017; 2021b; Haspelmath & Karjus 2017)

This has long been a goal of **general linguistics ("g-linguistics")** (e.g. Hawkins (ed.) 1988), but many linguists have given more prominence to other goals:

- describing mental grammars (generative grammar, cognitive grammar)
- explaining the possibility of language acquisition ("Plato's Problem")
- describing particular languages ("p-linguistics")
- explaining particular language structures on the basis of their histories
- inferring population history on the basis of linguistic reconstruction

Generative linguists have often talked about "universal grammar" (UG), but it is currently unclear how this relates to grammatical universals.

Some generativists still assume a very rich UG that can explain many universals (e.g. Cinque 1999; Baker 2015), but others assume that UG is very impoverished, perhaps containing only Merge and Agree

(see my blogpost on this: https://dlc.hypotheses.org/2481)

In traditional generative grammar (Chomsky 1965), there are both **architectural** and **substantive** elements in UG, and these are thought to explain universal properties of human languages, e.g.

architectural: syntax vs. morphology
                     (or structure building vs. vocabulary insertion)
           underlying vs. surface structure
           grammar vs. lexicon

           ...
substantive: noun, verb, adjective, adposition
           [person], [number], [tense]
           CP, TP, vP, VP, DP, NP, ...
           OT constraints (Faithfulness, Markedness)
           ...

These can be regarded as causal explanatory factors if they are innate and thus constrain possible mental grammars.

Four types of constraints (= explanatory factors) that may explain universals (Haspelmath 2019):

functional-adaptive constraints:	not everything has fitness
bicognitive-representational constraints:	not everything can be represented
mutational constraints:	not everything can arise through change
acquisitional constraints:	not everything can be acquired

I will argue that in the domain of argument coding and voice marking, there are some robust universals that can be explained on the basis of **coding efficiency**, while no stable biocognitive-representational explanations have emerged (as far as I can see).

2. Variable coding: Passive voice, dative alternation, differential accusative

passive voice alternation:

(1) Yucatec Maya (Lehmann 2015: 1437, 1448)
    a. *t-u=méek'-ah*   *u   chaan xipbal   le   maamah-o'*
       PFV-3.SBJ=hug-CMPL   her little boy   the mother-DET
       'The mother hugged her little boy.'

    b. *h=méek'-**ab**   le chaan xipbal   **tuméen** u   maamah-o'*
       PFV=hug-PASS   the little boy   by   his mother-DET
       'The little boy was hugged by his mother.'

dative alternation:

(2) Mandarin Chinese (Li & Thompson 1981: 376)
    a. *Wǒ sòng-le tā yi píng jiǔ.*
       I   give-PFV 3SG   one bottle wine
       'I gave him a bottle of wine.'

    b. *Wǒ sòng-le yi píng jiǔ* **gěi** *tā.*
       I   give-PFV one   bottle wine to   3SG
       'I gave a bottle of wine to him.'

differential accusative marking (DOM):

(3) Purepecha (Mexico; Capistrán-Garza 2015: 31)
    a. (indefinite P)
       xuchá arhá-s-ka kurúcha
       we   ingest-PRF-1.IND   fish
       'We ate fish.'

    b. (definite P)
       *xuchá arhá-s-ka kurúcha-**ni***
       we   ingest-PRF-1.IND   fish-OBJ
   'We ate the fish.'

Claims:

    – what these three construction types share is **asymmetrical coding**
    (– asymmetric coding is a key notion for formulating a wide range of **universals**;
                                               cf. Haspelmath 2021b)

    – these universals can be explained through **functional adaptation:**

        specifically, as following from a **tradeoff** between
            – the tendency to minimize speaker effort
                     – leading to a preference for short coding –
            – and the tendency to maximize the effect on the hearer
                     – leading to a preference for robust coding.

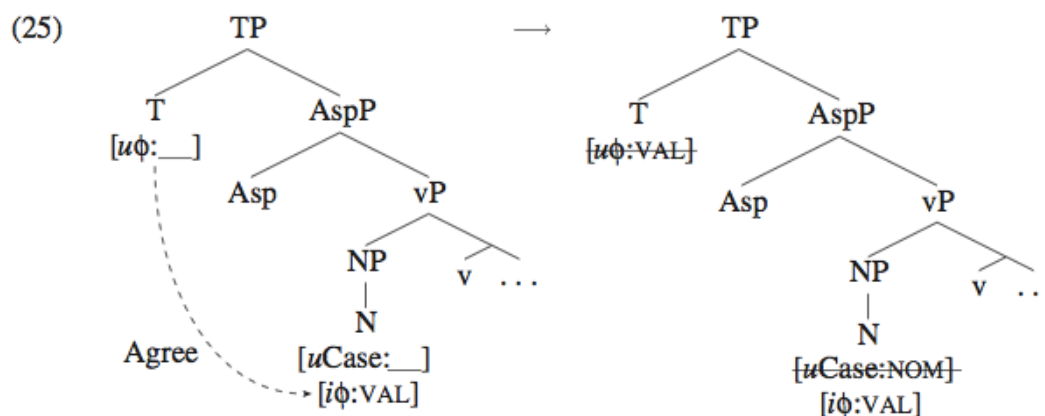## 3. Briefly on architectural approaches

One hopes that UG-based descriptions will eventually generalize across languages and across constructions.

For example, Larson (1988) tries to use the same formal machinery for the English Dative alternation and the English passive

(25)
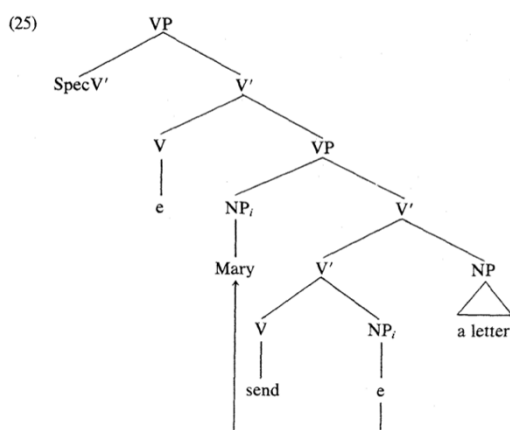
It should also be observed that the connection drawn here between Passive and Dative Shift is quite similar to that made within the framework of Relational Grammar (see Perlmutter (1983) and Perlmutter and Rosen (1984)). In Relational Grammar both Passive and Dative Shift are standardly viewed as instances of a single operation of *advancement*, which promotes argument phrases with respect to their grammatical relations. Thus, Passive is viewed as "2 → 1 advancement" and Dative Shift as "3 → 2 advancement," where "1," "2," and "3" designate the subject, direct object, and indirect object relations, respectively. Arguments that are supplanted in their grammatical

Kalin (2018), on differential object marking:

(25)

4. Briefly on semantic and information-structural approaches

These approaches tend to stay at the language-particular level:

• semantics:

One focuses on the language-particular details of the coding alternatives from a semantic point of view, e.g.

> – Goldberg (1995) on the semantic details of the English Dative alternation
> – von Heusinger & Kornfilt (2005) on the semantic details of Turkish DOM

• information structure:

One focuses on the information-structural differences between the coding alternatives, e.g.
- – Givón (1984) on the topicalization function of the dative alternation
- – Foley & Van Valin (1984) on the information-structural functions of passives
- – Dalrymple & Nikolaeva (2011), Iemmolo (2011) on the information-structural conditions for differential object marking

But these works generally stay at the language-particular level, and why they may show **good design** (they are "functionalist"), they are insufficient to demonstrate **adaptation** – they are not "functional-adaptive".

To demonstrate adaptation, one needs a comparative approach, involving testable universal claims.

> (This is what I mean by "Explaining X" – explanation at the g-level, going beyond p-description/p-explanation.)

5. Strong universal tendencies of asymmetrical coding

What passives, applicatives, dative alternation and DOM share: the coding is ASYMMETRIC.

agent-oblique alternation:

an alternation between a transitive construction (the "active"), and a construction in which the P of the active is coded like the intransitive S and the A of the active is flagged as an oblique (the "passive-intransitive")

agent(A)	patient(P)	verb	
	patient(S)	verb(-PASS)	agent(OBL)

stereotypical: | mother | hugs | boy | |
| boy | hug(-**PASS**) | **by**+mother |

dative alternation:

an alternation between a ditransitive construction with neutral alignment and a construction with indirective alignment

agent(A)	recipient(=P)	theme(=P)	verb
agent(A)	recipient (not = P)	theme (=P)	verb

stereotypical: | boy gives | mother | flower |
| boy gives | **to**+mother | flower |

differential object marking:

a situation where some kinds of object arguments get different coding from other kinds of object arguments

> stereotypical: I saw house-Ø
> I saw woman-**ACC**

Universals of coding:

U1: If there is special verbal marking in an agent-oblique alternation, it is found on the **passive** verb, and if there is special flagging, it is found on the **agent**.

U2: If there is special marking in a dative alternation, it is found on the **R-argument**.

U3: If there is special marking in a DOM construction, it is found on the **animate/definite/topical** argument.

Agent-oblique alternations:

Of course, most agent-oblique alternations have special marking on the verb (they are **passives**), but some languages have uncoded agent-oblique alternations:

(4) Bambara (Mande; Cobbinah & Lüpke 2012: 136)
>   a. *ù   bɛ   ɲɔ`   dan*
>   they   PRS   millet   sow
>   'They sow millet.'
>
>   b. *ɲɔ`    bɛ    dan    (u    fɛ`)*
>   millet   PRS   sow   they   by
>   'Millet is sown (by them).'

And of course, most passives have special marking on the oblique agent, but again this is not definitional – the oblique flag need not be longer than the subject flag, cf. the hypothetical:

> they-ERG sow millet
> millet sow-PASS they-INS

But whenever the coding is asymmetric, the oblique-agent flag is longer than the subject flag.

Dative alternations:

Stereotypical dative alternations of the English and Chinese type have only a dative flag and no other flags, so of course they have a special flag on the R argument. This kind of alternation is also found elsewhere (though not very commonly, cf. Siewierska 1998).

(5) Emai (Benue-Congo; Schaefer & Egbokhare 2010: 129)
>   a. *àlèkè   háé   ólí   ọmọ́hé   òsà*
>   Aleke   pay   the   man   debt
>   'Aleke repaid the man her debt.'

b. *àlèkè   háé   òsà   **lí**   ólí   ómóhé*
   Aleke   pay   debt   to   the   man
   'Aleke repaid her debt to the man.'

(6) Thai (Thepkanjana 2010: 415)
   a. *sǒmchaay   khǎay   rót   phʉ̌an*
      Somchaay   sell   car   friend
      'Somchaay sold a car to his friend.'

   b. *sǒmchaay   khǎay   rót   **kɛ̀ɛ** phʉ̌an*
      Somchaay   sell   car   to   friend
      'Somchaay sold a car to his friend.'

But this is not definitional – the dative flag need not be longer than the accusative flag, cf. the hypothetical:

        boy gives mother-ACC flower-ACC
        boy gives mother-DAT flower-ACC

And indeed, one sometimes finds the dative alternation in languages that have an accusative marker, e.g.

(7) Modern Standard Arabic (Ryding 2011: 290-291)
   a. *ʔaʕṭay-tu   l-bint-a   l-miftaaħ-a*
      give.PRF-1SG   DEF-girl-ACC   DEF-girl-ACC
      'I gave the girl the key.'

   b. *ʔaʕṭay-tu   l-miftaaħ-a   **li**-l-bint-i*
      give.PRF-1SG   DEF-key-ACC   to-DEF-girl-GEN
      'I gave the key to the girl.'

But whenever the coding is asymmetric, it is the R-argument that has the longer flag.

Differential object marking:
In almost all cases, special flagging means that the animate/definite/topical object has a flag, and the less referentially prominent argument lacks a flag.

(8)   Persian (Dalrymple & Nikolaeva 2011: 108-112)
   a. *man   ketâb-**râ**   xarid-am.*
      I   book-ACC   buy.PST-1SG
      'I bought the book.'

   b. *man   sib-i(*-râ)   xord-am.*
      I   apple-INDF(-ACC)   eat.PST-1SG
      'I saw an apple.' (accusative flag is not allowed on nontopical P)

   c. *ki   mašin-i-*(**râ**)   did?*
      who   car-INDF-(ACC)   see.PST[3SG]
      'Who saw a car?' (accusative flag is required on topical P)

But again, this is not definitional – a language could have two different accusative forms, one that is longer and another one that is shorter

        I saw house-ACC.SHORT
        I saw woman-ACC.LONG

Cf. German   *Ich saw de-**n** Amethyst.*   'I saw the amethyst.'   (inanimate)
                    *Ich sah de-**n** Analyst-**en***   'I saw the analyst.'   (animate)

6. The universal coding asymmetries correspond to universal frequency asymmetries:
short coding is used in the usual situations

– oblique-agent constructions are less frequent than "actives"
– dative alternants are less frequent than neutral-alignment patterns
– differentially marked objects are less frequent than unmarked objects

---

(9) **U4: The grammatical form-frequency correspondence universal**
     When two grammatical construction types that differ minimally (i.e. that form a semantic opposition) occur with significantly different frequencies, ***the less frequent construction tends to be overtly coded*** (or coded with more segments), while the more frequent construction tends to be zero-coded (or coded with fewer segments).

This can explained as following from a tradeoff between
    – the **tendency to minimize speaker effort** (leading to a preference for short coding), and
    – the tendency to maximize **clarity for the hearer** (leading to a preference for robust coding).
    More frequent meanings are more predictable and hence need less coding
    (see Haspelmath 2021b).

The form-frequency correspondence hypothesis makes a large number of correct predictions in various domains of grammar, but what is crucial for transitive and ditransitive constructions is the finding that **role rank** and **referential prominence** are generally associated:

---

(10) **U5: Usual role-reference associations** (Haspelmath 2021a)
     Arguments with higher-ranked roles tend to be more referentially prominent,
     and vice versa.

          (role rank: A > P, R > T)

(11) **referential prominence**
   a. inherent prominence
       person scale: locuphoric (1st/2nd) > aliophoric (3rd person)
       (full) nominality scale: person form (independent or index) > full nominal
       animacy scale: human (> animal) > inanimate

   b. discourse prominence
       specificity scale: definite (> specific indefinite) > indefinite nonspecific
       givenness scale: discourse-given > discourse-new
       focus scale: background > focus

When a clause type deviates from the usual associations, it is likely to get special coding:

(12) **U6: The role-reference association universal** (Haspelmath 2021a)
Deviations from usual associations of role rank and referential prominence tend to be coded by longer grammatical forms.

The coding universals that we saw earlier are special cases of this, just as U6 is a special case of U4 (the grammatical form-frequency correspondence universal).

7. Oblique-agent constructions and dative constructions occur with usual role-reference associations

It has often been observed that passives and dative alternants tend to be used when the patient is topical / when the recipient is not topical – i.e. when the argument roles do not have their usual referential-prominence values. I claim that this is a universal effect:

(13) **U7: Givenness of P in passives and other oblique-agent constructions**
If an oblique agent alternation is sensitive to givenness, then the oblique-agent alternant tends to be used when the A is not given information and/or the P is not new information.

stereotypical:	mother	hugs	boy
	boy	hug-**PASS**	**by**+mother

(14) **U8: Nongivennes of R in dative alternants**
If a dative alternation is sensitive to givenness, then the dative alternant tends to be used when the R is not given information and/or the T is not new information.

stereotypical: boy gives mother flower
                boy gives **to**+mother flower

U7 and U8 are *frequency universals* (special cases of U5), which explain *coding universals* such as U1-U3.

8. Parallels between passives and DOM in the very first work on differential object marking (Moravcsik 1978)

Moravcsik (1978) coined the term *differential object marking* (often attributed to Georg Bossong: 1985). She also gave the first formulation of what is universal:

"If there is any semantic difference between an accusative marking and … a nominative marking (to the exclusion of passivization), this semantic difference will be related … to definiteness, or to animacy, or humanness …, with the accusative … marking the more definite (rather than the less definite), [and] the animate or human (rather than the inanimate or non-human), … noun phrase." (Moravcsik 1978: 283)

In this statement, she excludes passivization, which she has to do because she works with the labels "nominative" and "accusative", rather than A and P (which also "exclude passivization" in a sense, see Haspelmath 2011).

But Moravcsik does note that there is a parallel between passivization and DOM:

> There are two distinct types of sentences in which objects appear marked as a transitive subject: in sentences which do not also include a transitive subject and in sentences which also include a transitive subject. The following examples illustrate these two sentence types.
>
> a. LATIN:
>
> puer puellam amat   "boy-nom. girl-accus. loves"
>   'The boy loves the girl.'
>
> puella a puero amatur   "girl-noml by boy-ablat. is-loved"
>   'The girl is loved by the boy.'
>
> puella amatur   "girl-nom. is-loved"
>   'The girl is loved.'
>
> puella beata est   "girl-nom. happy is"
>   'The girl is happy.'
>
> b. SPANISH:
>
> el chico ve la nieve   "the-nom. boy sees the-nom. snow"
>   'The boy sees the snow.'
>
> el chico ve a esa chica   "the-nom. boy sees accus. this girl"
>   'The boy sees this girl.'

And she also notes the parallel in topic status:

Both DOM and passivization (or more generally, the oblique-agent construction) tend to be used when the patient is topical, or more generally high in referential prominence (when there is a deviation from the usual role-reference associations)

Thus, Moravcsik (1978) was an extremely insightful article, and it was largely underappreciated.

9. Conclusion

We need a **functional-adaptive** approach for explanations at the g-level, and it must be universalist, because only universal tendencies can be explained in functional-adaptive terms.

The explanation proposed here is extremely general and involves highly **abstract** notions, such as "referential prominence" and the role-types A, P, R, T.

It seems to be supported by a large amount of cross-linguistic evidence, and the proposal by Moravcsik 81978) has proved robust – in contrast to generative approaches, where it is difficult to say what has proved robust.

But it does not make any reference to biocognitive-representational constraints – it invokes a highly general **efficiency** principle, and is indifferent with respect to architectural assumptions, such as

<div align="center">

word vs. phrase
grammar vs. lexicon
inflection vs. derivation

</div>

This does not mean that domain-specific innate categories are "rejected" (as in much functionalist work) – such categories simply play no role here.

It may well be that A and P, as well as R and T are instances of innate role-types in sime sense (maybe as innate cognitive "attractors").

One important aspect of the approach is that the **universalist explanation** is separate from **language-particular analyses** – comparison and description (or g-linguistics and p-linguistics) are distinct to a substantial extent.

This is important to highlight, because there are frequent misunderstandings (e.g. Himmelmann 2021; see Haspelmath 2020). Maybe much of universalist generative grammar can be seen as based on such a misunderstanding.

[Historical note: A closely related talk was presented in 2018 at the *Syntax of the World's Languages* conference in Paris; see Haspelmath 2018]

References

Baker, Mark C. 2015. *Case: Its principles and parameters*. Cambridge: Cambridge University Press.

Bossong, Georg. 1985. *Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr. (https://zenodo.org/record/4697660)

Capistrán-Garza, Alejandra. 2015. *Multiple object constructions in P'orhépecha: Argument realization and valence-affecting morphology*. Leiden: Brill.

Chomsky, Noam A. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic approach*. New York: Oxford University Press.

Cobbinah, Alexander & Friederike Lüpke. 2009. Not cut to fit: Zero coded passives in African languages. In Matthias Brenzinger & Anne-Maria Fehn (eds.), *Proceedings of the 6th World Congress of African Linguistics*, 153–165. Cologne: Köppe.

Crevels, Mily. 2010. Ditransitives in Itonama. In Andrej L. Malchukov, Martin Haspelmath & Bernard Comrie (eds.), *Studies in ditransitive constructions: A comparative handbook*, 678–709. Berlin: De Gruyter Mouton.

Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and information structure*. Cambridge: Cambridge University Press.

Essegbey, James. 2010. Inherent complement verbs and the basic double object construction in Gbe. In Enoch Oladé Aboh & James Essegbey (eds.), *Topics in Kwa syntax*, 177–193. Dordrecht: Springer. (20 August, 2014).

Foley, William A & Jr. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Givón, T. 1984. Direct object and dative shifting: Semantic and pragmatic case. In Frans Plank (ed.), *Objects: Towards a theory of grammatical relations*, 151–182. London: Academic Press.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: The University of Chicago Press.

Haspelmath, Martin & Karjus, Andres. 2017. Explaining asymmetries in number marking: Singulatives, pluratives, and usage frequency. *Linguistics* 55(6). 1213–1235. (doi:10.1515/ling-2017-0026)

Haspelmath, Martin. 2004. Explaining the Ditransitive Person-Role Constraint: A usage-based approach. *Constructions* 2. (doi:10.5281/zenodo.831408) (https://journals.linguisticsociety.org/elanguage/constructions/article/view/3073.html)

Haspelmath, Martin. 2011. On S, A, P, T, and R as comparative concepts for alignment typology. *Linguistic Typology* 15(3). 535–567.

Haspelmath, Martin. 2017. Explaining alienability contrasts in adpossessive constructions: Predictability vs. iconicity. *Zeitschrift für Sprachwissenschaft* 36(2). 193–231. (doi:10.1515/zfs-2017-0009)

Haspelmath, Martin. 2019. Can cross-linguistic regularities be explained by constraints on change? In Schmidtke-Bode, Karsten & Levshina, Natalia & Michaelis, Susanne Maria & Seržant, Ilja A. (eds.), *Competing explanations in linguistic typology*, 1–23. Berlin: Language Science Press. (http://langsci-press.org/catalog/book/220)

Haspelmath, Martin. 2020. The structural uniqueness of languages and the value of comparison for description. *Asian Languages and Linguistics* 1(2). 346–366. (doi:10.1075/alal.20032.has)

Haspelmath, Martin. 2021a. Role-reference associations and the explanation of argument coding splits. *Linguistics* 59(1). 123–174. (doi:10.1515/ling-2020-0252)

Haspelmath, Martin. 2021b. Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics* 57(3). 605–633. (doi:10.1017/S0022226720000535)

Hawkins, John A. (ed.). 1988. *Explaining language universals*. Oxford: Blackwells. (doi:10.5281/zenodo.7148287) (Accessed October 14, 2022.)

Himmelmann, Nikolaus P. 2021. Against trivializing language description (and comparison). *Studies in Language*. (doi:10.1075/sl.19090.him) (https://ling.auf.net/lingbuzz/005705) (Accessed January 18, 2022.)

Iemmolo, Giorgio. 2011. Towards a typological study of differential object marking and differential object indexation. University of Pavia PhD dissertation.

Jukes, Anthony. 2006. Makassarese (basa Mangkasara'): A description of an Austronesian language of South Sulawesi. University of Melbourne Ph.D. dissertation.

Kalin, Laura. 2018. Licensing and differential object marking: The view from Neo-Aramaic. *Syntax* 21(2). 112–159. doi:10.1111/synt.12153.

Larson, Richard K. 1988. On the double object construction. *Linguistic Inquiry* 19(3). 335–391.

Lehmann, Christian. 2015. Valency classes in Yucatec Maya. In Andrej L. Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages: A comparative handbook*, vol. 2, 1427–1480. Berlin: De Gruyter Mouton.

Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Merlan, Francesca. 1982. *Mangarayi*. Amsterdam: North-Holland.

Moravcsik, Edith A. 1978. On the case marking of objects. In Greenberg, Joseph H (ed.), *Universals of human language*, vol. 4: Syntax, 249–289. Stanford: Stanford University Press. (https://zenodo.org/record/4688136)

Ryding, Karin C. 2011. *A reference grammar of modern standard Arabic*. Cambridge: Cambridge University Press.

Schaefer, Ronald P. & Francis O. Egbokhare. 2010. On Emai ditransitive constructions. In Andrej Malchukov, Martin Haspelmath & Bernard Comrie (eds.), *Studies in ditransitive constructions*, 115–144. Berlin: Mouton de Gruyter.

Siewierska, Anna. 1998. Languages with and without objects: The Functional Grammar approach. *Languages in Contrast* 1(2). 173–190.

Thepkanjana, Kingkarn. 2010. Ditransitive constructions in Thai. In Andrej Malchukov, Martin Haspelmath & Bernard Comrie (eds.), *Studies in Ditransitive Constructions*, 409–426. Berlin: De Gruyter Mouton.

Tournadre, Nicolas. 1995. Tibetan ergativity and the trajectory model. In Y. Nishi, James A Matisoff & Yasuhiko Nagano (eds.), *New horizons in Tibeto-Burman morphosyntax*, 261–275. Osaka: National Museum of Ethnology.

von Heusinger, Klaus & Jaklin Kornfilt. 2005. The case of the direct object in Turkish: Semantics, syntax and morphology. *Turkic Languages* 9. 3–44.