

How to present data in a paper

When presenting data in blocks or datasets in a linguistics paper (or, for that matter, presenting definitions or lists in inset blocks), it is important to keep in mind two basic principles:

- a. the data must be presented first, as soon as possible after its first mention, *before* it is discussed in detail;
- b. the data must be *introduced* in a general way alerting the reader as to why it is there, what it is intended to illustrate, and what they should look for in the dataset, and it should be immediately *explicated* afterwards, highlighting the specific details in the dataset that are of interest, and how these contribute to the paper.

The first principle ensures that readers do not have to flip back and forth between the data and the text, and that they do not have to follow a discussion of something hitherto unseen. The second principle prepares the reader ahead of time for what is coming and makes it easy to find the salient features of a dataset, and afterwards ensures that readers understand the examples and how you are interpreting them for the purposes of your writing.

An example for this type of writing might be drawn from a discussion of causative constructions, where the aim is to illustrate the morphological causative and introduce to the reader to the idea that there are some typological problems with the traditional distinction drawn between the semantic roles of AGENT and the CAUSER. We would want to begin this discussion by giving an example of a simple verb, followed by a sentence containing its causative derivative, as in (1):

Upper Necaxa Totonac (Totonac-Tepehua family)

- (1) a. kit na-ik-**skúx**-a čur:wá watsá
I FUT-1SG.SUB-work-IMPF now here
'I'll work here now'
- b. li:-lqʔapu:ci: akšní iš-kin-ta-**ma:skux-ú:** a:xcananú
INST-be.sad when PAST-1OBJ-3PL.SUB-CS-work-CS back.then
'it was sad when they made me work back then' (author's fieldnotes)

(1) contrasts the monovalent verb *skux-* 'work' with its causative derivative, *ma:skuxú:* 'make sby work', which is bivalent. The new semantic actant in (1b) becomes the subject, and the "displaced" subject of the base — generally referred to as the CAUSEE — is realized as a direct object. The new semantic actant has the role of CAUSER, which differs from AGENT by virtue of being the initiator of an unspecified event that in turn triggers the event expressed by the verbal base (Langacker 1987). In languages with morphological causatives, verbs whose subjects express prototypical AGENTS tend to be underived stems whereas those whose subjects are clearly CAUSERS tend to be derived. However, both within and across languages there is a certain fuzziness as to where the line between the two roles is drawn, and which predicates are derived or underived.

How NOT to present data in a paper

Do NOT discuss data that you haven't presented yet

WRONG

Upper Necaxa Totonac displays some apparent irregularities in the expression of subject and object agreement in cases where the subject is first- or second-person and the object is second- or first-person, and one or both is plural, as seen in (2). These forms are three-way ambiguous; the 1 > 2 form in (2a), for instance, which consists of the first-person subject prefix *ik-*, the plural object marker *ka-*, and the second-person object suffix *-n*, has not only the expected 1SG > 2PL reading, it is also used in situations where 1PL > 2PL and 1PL > 2SG. Likewise, the form in (2b) has three readings, 2SG > 1PL, 2PL > 1PL, 2PL > 1SG, although in this case it consists of a non-compositional configuration of affixes, the first-person object prefix *kin-*, the first-person plural subject suffix *-w*, and the reciprocal prefix *la-*.

- (2) a. *ika:túksni*
ik-ka-tuks-ni
1SG.SUB-PL.OBJ-hit-2OBJ:PFV
'I hit you guys'
'we hit you guys'
'we hit you'
- b. *kila:túkswi*
kin-la-tuks-wi
INST-RCP-hit-1PL.SUB:PFV
'you hit us'
'you guys hit us'
'you guys hit me'

RIGHT

Upper Necaxa Totonac displays some apparent irregularities in the expression of subject and object agreement in cases where the subject is first- or second-person and the object is second- or first-person, and one or both is plural. These forms are three-way ambiguous, as seen in (3), which shows a verb expressing action of the first-person on the second (1 > 2):

- (3) *ika:túksni*
ik-ka-tuks-ni
1SG.SUB-PL.OBJ-hit-2OBJ:PFV
'I hit you guys'
'we hit you guys'
'we hit you'

The form in (3) consists of the first-person subject prefix *ik-*, the plural object marker *kar-*, and the second-person object suffix *-n*. This is the form expected for the 1SG > 2PL reading, but it is also used in situations where 1PL > 2PL and 1PL > 2SG.

In a similar vein, the form in (4) expresses action of 2 on 1 and has three readings, 2SG > 1PL, 2PL > 1PL, 2PL > 1SG:

- (4) kila:túkswi
 kin-lá:-túks-wi
 INST-RCP-hit-1PL.SUB:PFV
 ‘you hit us’
 ‘you guys hit us’
 ‘you guys hit me’

Unlike (3), however, this form consists of a non-compositional configuration of affixes—the first-person object prefix *kin-*, the first-person plural subject suffix *-w*, and the reciprocal prefix *la:-*.

Do NOT concentrate all your data into a single long set then discuss it

WRONG

Upper Necaxa Totonac has a total of six valency-increasing morphemes, two causatives and three applicatives:

- (5) a. kima:ʔka:kní:ya
 kin-ma:-ʔka:k-ni:-ya
 1OBJ-CS-hot-CS-IMPf:2SG.SUB
 ‘you are making me hot’
- b. maʔajikwánli tzamá ʔawáča
 maʔa-jikwán-li tzamá ʔawáča
 STM-afraid-PFV that boy
 ‘he frightened the boy’
- c. na:kšʔoniyá:n ištapał
 na-ik-šʔó-ni-ya:-n išt-apał
 FUT-1SG.SUB-pay-BEN-IMPf-2OBJ 3PO-boy
 ‘I’ll pay you its price’

- d. šu:nú:k nali:ʔe:chi:kán ša:
 šu:nú:k na-li:-ʔe:-chi:-kan ša:
 bark FUT-INST-back-tie-IDF sweatlodge
 ‘they will tie it onto the sweatlodge with bark’
- e. nakinta:pína
 na-kin-ta:-pin-a
 FUT-1OBJ-CMT-go:2SUB-IMPF:2SG.SUB
 ‘you will go with me’
- f. talaʔáʔ ja: ištama:ki:ní: išlu:wakán
 ta-laʔ-ʔan-ʔ ja: iš-ta-ma:ki:-ní: iš-lú:wá-kan
 3PL.SUB-ALT-go-PFV where PAST-3PL.SUB-keep-PF 3PO-snake-PL.PO
 ‘they went to where they kept their snake’

The first two valency-increasers are causatives. (5a) shows the cross-linguistically typical causative that adds a volitional agent to the event expressed by the verb; (5b) shows the stimulus morpheme *mʔa-*, which typically adds a non-human or inanimate causer, or is used to form verbs expressing caused internal or psychological processes. (5c) illustrates the benefactive suffix, *-ní*, which adds a beneficiary, maleficiary, recipient, or experiencer to an event. The instrumental applicative, *li-*, shown in (5d), typically adds an instrument, but can also add motives. The comitative *ta-* seen in (5e) adds a co-actor that performs the action described by the verb along with the subject. The final applicative, the allative *laʔ-*, adds a goal to a limited set of motion verbs, mostly those based on the verbs *ʔn* ‘go’ and *min* ‘come’.

RIGHT

Upper Necaxa Totonac has a total of six valency-increasing morphemes, two causatives and three applicatives. The causatives are *ma-* *-ni:* ‘causative’ (6a) and *mʔa-* ‘stimulus’ (6b):

- (6) a. kima:ʔka:kní:ya
 kin-ma:-ʔka:k-ni:-ya
 1OBJ-CS-hot-CS-IMPF:2SG.SUB
 ‘you are making me hot’
- b. mʔaʔajikwánlĭ tzámá ʔawáča
 mʔa-jikwán-lĭ tzámá ʔawáča
 STM-afraid-PFV that boy
 ‘he frightened the boy’

Of the two, *ma-* *-ni:* is the cross-linguistically typical causative that adds a volitional agent to the event expressed by the verb; the stimulus morpheme *mʔa-*, on the other hand, typically adds

a non-human or inanimate causer, or is used to form verbs expressing caused internal or psychological processes.

The first of the four applicatives is the benefactive suffix, *-ni* shown in (7):

- (7) *na:kʃoʔoniyá:n iʃtapáʔ*
 na-*ik*-*ʃoʔó*-*ni*-ya:-n *iʃ*-tapáʔ
 FUT-1SG.SUB-pay-BEN-IMPF-2OBJ 3PO-boy
 ‘I’ll pay you its price’

This suffix adds a beneficiary, maleficiary, recipient, or experiencer to an event.

The example in (8) illustrates the instrumental applicative, *li:-*:

- (8) *ʃu:nú:k nali:ʔe:chi:kán ʃa:*
ʃu:nú:k na-li:-ʔe:-chi:-kan ʃa:
 bark FUT-INST-back-tie-IDF sweatlodge
 ‘they will tie it onto the sweatlodge with bark’

Not unexpectedly, this prefix typically adds an instrument to the event expressed by its base, but may also be used to add a motive for the event.

The comitative applicative adds a co-actor to the event:

- (9) *nakinta:pína*
 na-kin-ta:-pin-a
 FUT-1OBJ-CMT-go:2SUB-IMPF:2SG.SUB
 ‘you will go with me’

As shown in (9), the co-actor performs the action described by the verb along with the subject.

The final applicative, the allative *laʔ-*, adds a goal to an event:

- (10) *talaʔáʔ ja: iʃtama:ki:ní: iʃlu:wakán*
 ta-laʔ-*an*-*ʔ* *ja:* *iʃ*-ta-ma:ki:-ní: *iʃ*-lu:wá-kan
 3PL.SUB-ALTV-go-PFV where PAST-3PL.SUB-keep-PF 3PO-snake-PL.PO
 ‘they went to where they kept their snake’

The allative applicative combines with only a limited set of motion verbs, mostly those based on the verbs *an* ‘go’ and *min* ‘come’.

Do NOT just give a complicated example and the expect the reader to figure it out

WRONG

Lushootseed oblique-centred relative clauses like that in (11) require nominalization of the embedded verb:

(11) ǰʷul' čəd ɬuləʔuǰʷtxʷ tiʔəʔ ɬadsʔəɬtxʷ

ǰʷul' čəd ɬu=lə-ʔuǰʷ-txʷ tiʔəʔ ɬu=ad=s=ʔəɬ-txʷ
only 1SG.SUB IRR=PROG-go-ECS PROX IRR=2SG.PO=NM=eat-ECS
'I will just be taking [them] what you will feed [them] with'

(Hess 1998: 58, line 56)

As in all such relatives, the nominalized clause realizes its “subject” as a possessor.

RIGHT

Lushootseed oblique-centred relative clauses like that shown in the sentence in (12) require nominalization of the embedded verb, here ʔəɬtxʷ ‘feed sby with sth’:

(12) ǰʷul' čəd ɬuləʔuǰʷtxʷ tiʔəʔ ɬadsʔəɬtxʷ

ǰʷul' čəd ɬu=lə-ʔuǰʷ-txʷ tiʔəʔ ɬu=ad=s=ʔəɬ-txʷ
only 1SG.SUB IRR=PROG-go-ECS PROX IRR=2SG.PO=NM=eat-ECS
'I will just be taking [them] what you will feed [them] with'

(Hess 1998: 58, line 56)

In this example, the headless relative clause *tiʔəʔ ɬadsʔəɬtxʷ* ‘what you will feed them with’ refers to the food, the oblique object of the verb ʔəɬtxʷ ‘feed sby with sth’; because the referent of the clause is an oblique object, the clause is nominalized with the nominalizing clitic *s=* and second person is realized as a possessive prefix, *ad=*, rather than as a subject clitic.

Do NOT simply end a section or a topic with a dataset

WRONG

A very common error students (and some professionals) make, usually in conjunction with the “don’t discuss data you haven’t presented error,” is giving data or placing an inset and then simply moving on to the next topic:

Datasets (and insets) should always be followed by a wrap-up either of the data (an explanation of the example) or of the topic

RIGHT

A very common error students (and some professionals) make, usually in conjunction with the “don’t discuss data you haven’t presented error,” is giving data or placing an inset and then simply moving on to the next topic:

Datasets (and insets) should always be followed by a wrap-up either of the data (an explanation of the example) or of the topic

Avoiding this violation of principle (b) above ensures good continuity and flow in your paper, and adhering to it usually forces you to think carefully about whether, indeed, you have made it clear enough to the reader what the data they just saw in the first place was actually about. It also gives you a chance to add corollary information that might be helpful to the discussion, such as the observation that following this last rule avoids the error of ending a *paragraph* with a dataset as well (since paragraphs usually enjoy some topical unity). Good stylistic practice usually dictates that the paragraph following a dataset NOT be indented (in styles with indentation on the first line of every paragraph).