Linguist's Assistant: A Resource For Linguists

Stephen Beale
University of Maryland, Baltimore County
Baltimore, MD
sbeale@cs.umbc.edu

Tod Allman
Onyx, Consulting
Baltimore, MD
todallman@yahoo.com

Abstract

The Linguist's Assistant (LA) is a practical computational paradigm for describing languages. In this paper we describe how to use LA with naturally occurring texts that exemplify interesting target-language linguistic phenomena. We will describe how such texts can be semantically analyzed using a convenient semi-automatic document authoring interface, in effect adding them to LA's standard semantic-based elicitation corpus. We then exemplify the language description process using a phenomenon that is prevalent in our research: alienable vs. inalienable nominal possession.

1 Introduction

The Linguist's Assistant (LA) is a practical computational paradigm for efficiently and thoroughly describing languages. Previously (Beale, submitted) we reported on the first of three main modes of LA-based language description: using the provided elicitation corpus of semantically analyzed sentences as the starting point and organizing principle from which the user describes the linguistic surface forms of a language using LA's visual lexicon and grammatical rule development interface. We described the semantic representation system that we developed and the make-up of the corpus of semantically analyzed texts that are meant to provide examples of a large subset of the kinds of meaning found in written communication. We described the visual lexicon and grammatical rule development interface that the linguist uses to record the lexical and grammatical knowledge needed to translate the semantic corpus into the target language.

Having read this previous paper, a respected linguist offered some valid criticism (valid, that is, if LA were restricted to this first mode of operation): "I am, in general, a bit reluctant to use ready-made questionnaires, for all sorts of reasons -- some of which you mention yourself. It so happens that my personal interest has always been on naturalistic speech… I have always paid a lot of attention to what actually shows up in everyday spoken speech, as opposed to what could exist 'grammatically' but is never heard. I've always wondered why so many grammars or articles in linguistics work on sentences such as '*The man sees the woman.*' which don't appear ever in naturalistic speech." (Alex François, personal communication). This paper is an attempt to counter such criticism by describing the second mode of operation in LA-based language description: acquiring language data and grammatical knowledge using naturally occurring texts that exemplify interesting target-language linguistic phenomena. We will describe how such texts can be semantically analyzed using a convenient semi-automatic document authoring interface, in effect adding them to the standard semantic-based elicitation corpus used in the first mode of operation. We exemplify the process using a linguistic phenomenon that is prevalent in Oceanic languages: alienable vs. inalienable nominal possession. Ishizuka (2010) describes a similar phenomenon in Japanese and Korean.

Before moving on, we should mention the third mode of LA-based language description: acquiring knowledge (lexical and grammatical) to cover pre-authored stories ("authored" in our context means that a semantic representation has been prepared). The semantically motivated elicitations from mode one combined with knowledge gained from the naturally occurring texts of mode two provide a solid foundation for lexicon and grammar development, but we have found that adding to that the experience and discipline of acquiring the knowledge necessary to generate actual story-length texts is invaluable. This is usually the best opportunity for documenting phenomena that is more lexically de-

Proceedings of the 9th Workshop on Asian Language Resources, pages 41–49,
Chiang Mai, Thailand, November 12 and 13, 2011.

pendent since the vocabulary in the semantic-based elicitation stage is quite limited. It also provides a test bed for the knowledge acquired in the first two modes of operation. For this reason we include several pre-authored community development texts/stories with LA. After acquiring the necessary lexical and grammatical knowledge for the target language, a draft translation of the stories can be produced and checked for naturalness and accuracy. LA has been used in this mode to produce a significant amount of high-quality translations in Jula (a Niger-Congo language), Kewa (Papua New Guinea), North Tanna (Vanuatu), Korean and English. Work continues in Vanuatu, with additional languages planned in the near future. We argue that the high quality results achieved in these translations demonstrate the quality and coverage of the underlying language description that LA produces. Beale et al. (2005) and Allman and Beale (2004; 2006) give more information on using LA in translation and for documentation on the evaluations of the translations produced.

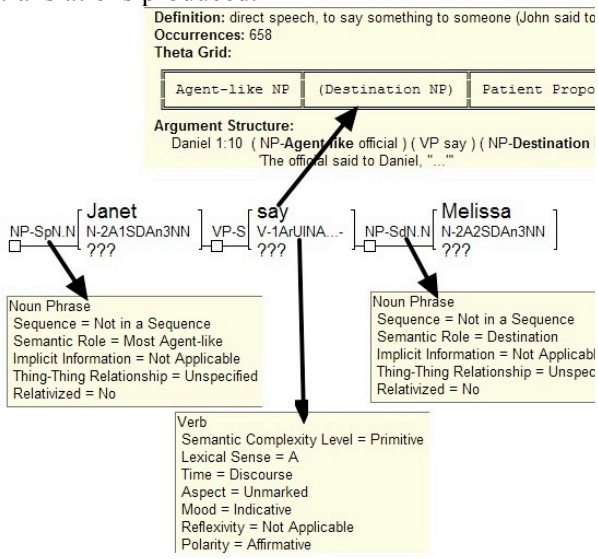


Figure 1: Semantic representation

LA is available for academic research and non-profit applications. Tutorials and related papers are also available, although a significant portion of our planned work is to produce better tutorials and workshop materials. The developers plan to offer tutorials at various conferences in the near future. We emphasize that LA is a work in progress. In any practical product with complex theoretical underpinnings there is a development loop where a "critical mass" of theory is implemented, the surrounding support tools created and tested, the product is used and evaluated, and then work begins again on improving the theoretical base. LA is somewhere in the late stages of the "being used and evaluated" step of this cycle. We certainly intend to improve the theoretical basis of each aspect of the product as time goes on, in large part as a result of the feedback, suggestions and criticisms of our users.

2 Introduction to LA

Consult Beale (submitted) for details on LA, including the semantic representation language and the visual lexicon and grammar interfaces. In order to make this paper self-contained, we summarize some of this material here.

A top-level view of the semantic representation is shown in Figure 1. Each concept "bundle" takes up three lines of text (for example, in the middle of the figure, "say", the line directly below that with "V-1ArUINA", and the line with the question marks). The top line is the English gloss of the concept. It must be emphasized that the English gloss is only shown for convenience; it represents a concept in our ontology. The upper yellow box appears upon a mouse-over of the concept; it provides details about the definition and usage of the concept. The middle line of the concept bundle consists of letters and numbers that specify the semantic features associated with this particular instance of the concept. These features can be viewed by placing the cursor over the concept. Note that phrases and clauses have semantic features defined for them in the same way that instances of concepts do. As shown in the figure, noun phrases have features such as semantic role. Verbs have time, aspect and mood features. Nouns have features that specify person, number and various reference-related meanings. The bottom line of the concept bundle is the "translation" of the concept into whatever target language is currently loaded (in the figure, question marks are displayed because no language is loaded). This "translation" is only a mapping to a target root word; often a concept will require more than a direct word-for-concept substitution.

Each sentence in the elicitation corpus has a semantic representation. In the first mode of language description described in Beale (submitted), the linguist needs to "teach" the computer how to realize each of the parts of that input semantic representation, including all the individual concepts, each of the semantic features and all the relationships (such as the case role relationships, discourse relations and adposition relationships). Backing up a bit, it is important to think about the overall nature of an LA project. The elicita-

tion corpus contains the wide range of phenomena that we are interested in documenting. The linguist creates the lexical knowledge and grammatical rules so that LA's built-in text generator can accurately translate the underlying meaning of the included semantic-based elicitation corpus. After this first stage is complete, the second stage of language description that is highlighted in this paper begins: using naturally occurring target language texts to describe important linguistic phenomena that occur in the language. Once the descriptive phases are complete, the resulting computational model of the language can be used in translation applications or output as part of a language documentation project.

	Stems	Glosses	type
1	aprend	learn	-er
2	habl	speak	-ar
3	ten	have	-er
4	viv	live	-ir

Figure 2: Lexical features for Spanish

	Stems	Glosses	infinitive	present indic 1st sing
1	aprend	learn	aprender	aprendo
2	habl	speak	hablar	hablo
3	ten	have	tener	tengo
4	viv	live	vivir	vivo

present indic 2nd sing	present indic 3rd sing	present indic 1st pl	present indic 3rd pl
aprendes	aprende	aprendemos	aprenden
hablas	habla	hablamos	hablan
tienes	tiene	tenemos	tienen
vives	vive	vivimos	viven

Figure 3: Lexical forms for Spanish

How does the linguist "teach the computer how to realize"? LA provides a rich, visual interface for building target lexicons and grammatical rules. Figure 2 shows the interface for creating and displaying lexical features, for example, the inflection type (-ar, -er or –ir) of a Spanish verb. Figure 3 shows the interface for displaying forms. Lexical form generation rules can be written to automatically generate each of the forms of a word. The white boxes in Figure 3 are irregular forms that were corrected by the user.

Grammatical rules typically describe how a given semantic structure is realized in the language. The whole gamut of linguistic phenomena is covered, from morphological alternations to case frame specifications to phrase structure ordering to lexical collocations – and many others. Figures 4-8 are examples of various types of grammatical rules. Figure 4 shows a morphophonemic rule; Figure 5 a phrase structure ordering rule; Figure 6 a feature copying rule (as would be used, for example, in Subject-Verb agreement in English), Figure 7 a table, and Figure 8 a theta-grid (or case-frame) realization rule. There are also rules similar to Figure 8 for converting the base semantic representation to a deep structure that is more appropriate for the target language. For example, Kewa[1] has a rule that converts the basic semantics for "X respects Y" into "X lifts-up the name of Y."

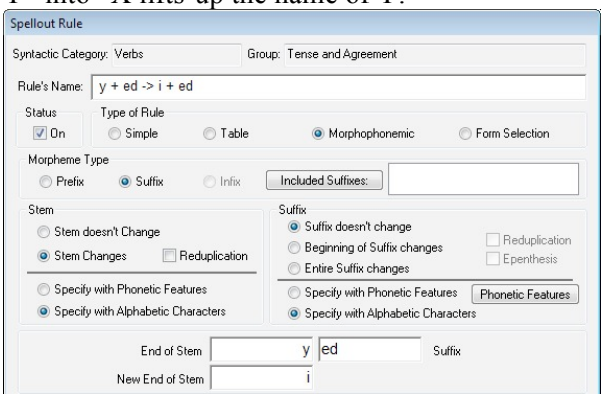


Figure 4: Morphophonemic rule

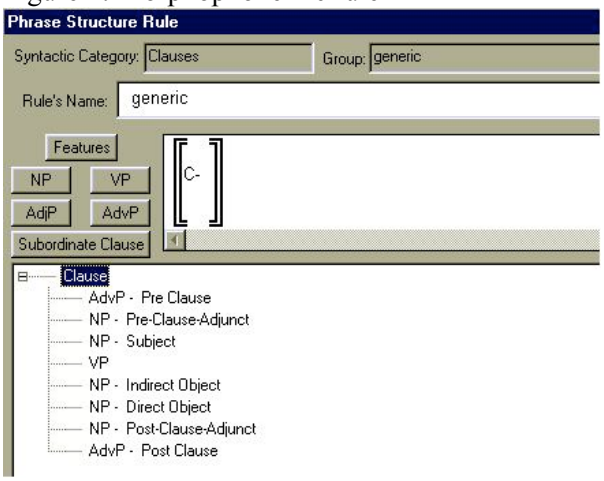


Figure 5: Phrase structure ordering rule

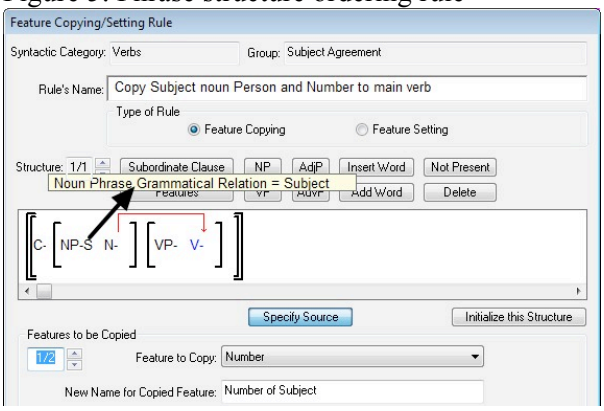


Figure 6: Feature copying rule

Currently, the linguist is responsible for the creation of rules, albeit with a natural, visual interface that often is able to set up the requisite input semantic structures automatically. We continue work on modules that will allow the semi-

[1] Dr. Karl Franklin supplied the Kewa data.

automatic generation of rules similar to research in the BOAS (McShane et al., 2002), LinGO (Bender et al., 2010), PAWS (Black and Black, 2009) and Avenue (Probst et al., 2003) projects. Such modules will, we believe, make LA accessible to a larger pool of linguists. We also provide a growing list of rule templates that linguists can use to describe common phenomena.

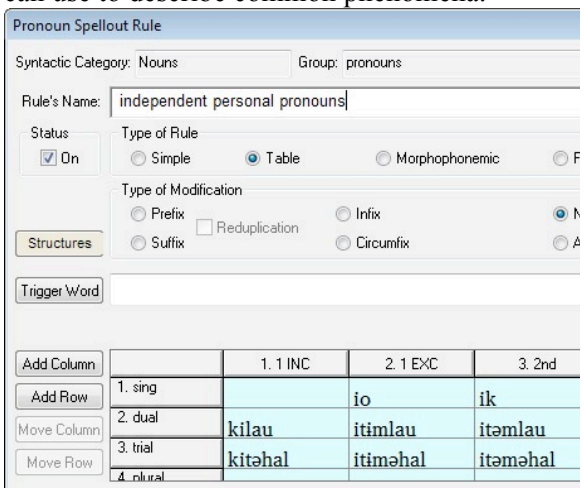


Figure 7: Table rule

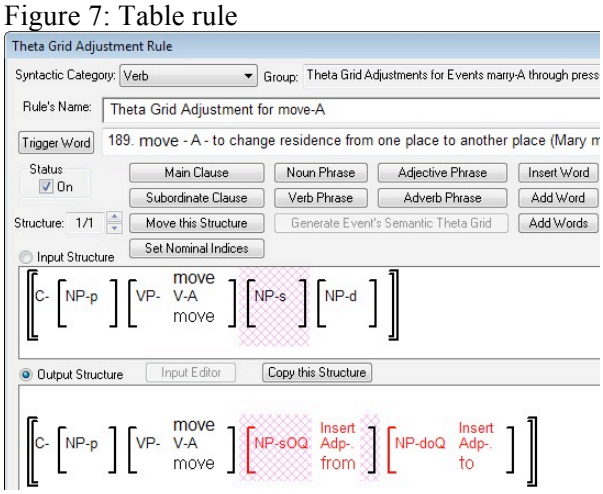


Figure 8: Theta grid (or case-frame) rule

3 Possession in Maskelynes

This paper focuses on using LA to describe a particular linguistic phenomenon using naturally occurring texts. We use alienable vs. inalienable nominal possession in Maskelynes as our case study. Ishizuka (2010) describes a similar phenomenon in Japanese and Korean. Maskeleynes is an Oceanic language spoken by about 1400 people in central Vanuatu. The language data and analysis presented here was inspired by a draft version of David Healey's doctoral thesis (which he has asked me not to directly reference in its present form). In addition to the draft nature of that document, some of the data presented below was extrapolated from the examples without confirmation by a native speaker. As such, we do not intend this to be a linguistic specification. However, the phenomenon described is typical of Oceanic languages and has been directly observed by the author in other Vanuatu languages. For the workshop presentation we will augment or replace this example with one more directly relevant to Asia; the point here is to describe the scope and methodology of LA as a linguistic resource.

Oceanic linguists have historically divided nouns into alienable and inalienable classes.[2] Inalienable nouns always appear with their "possessor." For example, body parts ("my arm") and relatives ("John's father") must occur with their possessor, as just illustrated. Healey also reports the more recently understood distinction of direct vs. indirect possession (Lichtenberk, 1985), which occurs in addition to the alienable vs. inalienable distinction. Directly possessed nouns in Maskelynes carry the marker of possession on the head noun whereas indirectly possessed nouns carry the possessive marker on the possessor noun or pronoun. We summarize the data in the rest of this section and in Figure 9.

In Maskelynes, inalienable nouns (section 1 of Figure 9) can either be directly possessed or indirectly possessed, depending on the class of the noun. Kinship terms and visible body parts generally are directly possessed (section 1A). Directly possessed inalienable nouns take an obligatory possession suffix. If the possessor must be specified (beyond the pro reference of the possession suffix), the possessor noun follows the head noun with no additional marking.

Maskelynes also has indirectly possessed inalienable nouns (section 1B). Some inanimate nouns that must be referred to with a possessor (for example, "his song" and "the home's shadow") and many internal body parts are indirectly possessed. These all follow the 'h' class of indirectly possessed nouns described below.

All alienable nouns (section 2) are indirectly possessed.

All indirectly possessed nouns (sections 1B, 2A and 2B) are either in the 'h' class (section 2A, typically foods and drinks) or the 's' class (section 2B, general nouns). Indirect possession can be realized with a possessive pronoun (that agrees with the 'h' or 's' class, as appropriate), or, when the possessor cannot be a pronoun, a genitive proclitic (hX- or sX- depending on the noun class) attached to the possessor noun, in

[2] See (Ishizuka 2010) for a treatment of alienable vs. inalienable possession in Japanese and Korean.

which case the nominaliser (which generally occurs on nouns) of the possessor is deleted.

1. Inalienable
 A. Directly possessed
  i. Human
   Pro-suffix possessor: *a-na-gw*
   NOM-mother-POSS.1st.excl.sing
   "my mother"
   Possessor must be specified: *a-na-n a-vanuan*
   NOM-mother-POSS.3rd.sing NOM-man
   "the man's mother"
  ii. Non-human
   Pro-suffix possessor: *nX-rie-gw*
   NOM-leg-POSS.1st.excl.sing
   "my leg"
   Possessor must be specified: *nX-rie-n a-vanuan*
   NOM-leg-POSS.3rd.sing NOM-man
   "the man's leg"
 B. Indirectly possessed
  Pronomial possessor: *nX-bwe hagw*
  NOM-song POSSPRO.1st.excl.sing
  "my song"
  Possessor must be specified: *nX-bwe hX-vanuan*
  NOM-song POSS.1st.excl.sing-man
  "the man's song"
2. Alienable (and indirectly possessed)
 A. h class
  Pronomial possessor: *nX-buai hagw*
  NOM-pig POSSPRO.1st.excl.sing
  "my pig"
  Possessor must be specified: *nX-buai hX-vanuan*
  NOM-pig POSS.1st.excl.sing-man
  "the man's pig"
 B. s class
  Pronomial possessor: *nX-kuvkuv sagw*
  NOM-axe POSSPRO.1st.excl.sing
  "my axe"
  Possessor must be specified: *nX-kuvkuv sX-vanuan*
  NOM-axe POSS.1st.excl.sing-man
  "the man's axe"

Figure 9: Possession Examples for Maskelynes

A final noun class relevant to our discussion involves the nominaliser: the human vs. non-human class. Nouns in the human class take the 'a-' nominaliser (section 1Ai) whereas non-human nouns (section 1Aii) take the 'nX-' nominaliser. The 'X' in this nominaliser and in the proclitic is phonologically conditioned; we will leave it as 'X' to simplify the discussion.

The examples in Figure 9 give an exhaustive reckoning of the different realization possibilities (other than the fact that different persons and numbers can be used for the pronouns and possessive suffixes).

4 Describing Possession Using LA

4.1 Authoring Examples

The first step in describing a new phenomenon in LA is to author examples. This process will produce semantically analyzed examples that will be used in the knowledge acquisition stage (sections 4.2 through 4.4 below) and in the testing stage (section 4.5).

Upon starting the document author, the system will ask for input sentences in a controlled English.[3] This is a key benefit and a limitation at the same time. LA is not able to parse target texts into the semantic representation that is needed in the subsequent stages (since it does not have a complete target grammar at this stage). Therefore we allow the user to mentally translate the meaning of the target language examples into the restricted English. Our built-in English analyzer will then semi-automatically produce the semantic analysis with only a small amount of editing of the results required from the user. Of course it would be optimal if the user could enter the examples in the target language and have an automatic semantic analysis performed, but this is impossible in the absence of a target language analyzer. Work has begun, however, on tools that will allow computer-assisted semantic analysis of target texts, which would obviate the need for the user entering simplified English translations.

To describe possession in Maskelynes, the user first enters the restricted English translations of the target sentences from Figure 9, as shown in Figure 10. In practice, because the authoring stage is relatively simple, the user could enter many more examples than shown, including an exhaustive accounting of all the possible combinations of number and person of both the head and possessor nouns. Figure 11 shows (in admittedly small print) a version of the results of the built-in semantic analysis. In the case of the first input phrase "My(John's) mother", the analysis is correct and no further editing is required. Notice that the analyzer chose the correct number and person (1st singular) for "John" and the correct Kinship relationship. For "my leg" the analyzer correctly chose the Body-Part relationship. We continue to work on the

[3] See Beale et. al (2005) for a description of the controlled English and for a description of the authoring process.

accuracy of the built-in English analyzer, but minor adjustments are sometimes necessary.

4.2 Setting Up Target Language Features

The key to describing possession in LA for Maskelynes - and indeed the key to using LA in general - is to first identify and create the target language features that will make it easy to write surface rules. In the case of possession, we need to know what class of noun is involved: inalienable vs. alienable, direct vs. indirect, and human vs. non-human. The user can define these features in the LA lexicon and specify the correct value for each for each noun root. Figure 12 shows an example lexicon for Maskelynes' nouns. Note the "class" and "human?" features.

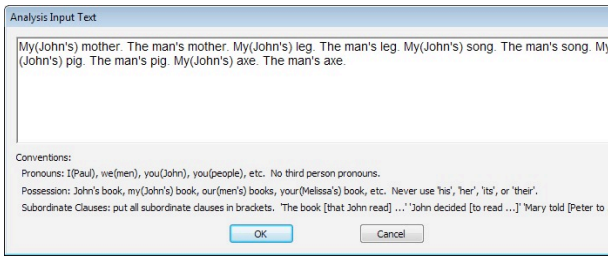


Figure 10: Document authoring input text

As you can see in Figure 9, it is also necessary to know whether the possessor noun is specified or whether there can be a suffix or a full pronoun reference to it. A key step in describing possession is to define such a feature on nouns, "Realization Type," which takes the values "noun" or "pronoun." The user can delay consideration of how to actually set that feature for later. To conserve space we do not show the trivial process of defining a new target feature.

Some words in Figure 9 have a POSS suffix (like *a-na-gw* in 1Ai). Some words have a gentive proclitic (like *hX-vanuan* in 2A). And other words have no possession affixes (like *nX-bwe* in 2A). Therefore another target feature that will make the final surface production rules easier is the "Noun Possession Type" feature for nouns. This will take the values "none", "possessed" and "genitive proclitic". This is the key feature used in our discussion in section 4.3 below. Again, the assignment of the correct value to this feature can be assumed when writing the rule that generates the actual surface form (section 4.3); later (section 4.4) the rule(s) will be written to set the correct value.

Note the difference between lexical features that are associated with a given root (and are defined in the lexicon) and the general target features that are defined outside the lexicon and whose values must be set by some rule. Given the two general features ("Realization Type" and "Noun Possession Type") and the two lexical features (class and human?), it will be possible to construct surface rules that implement the full range of possible realizations of possession.

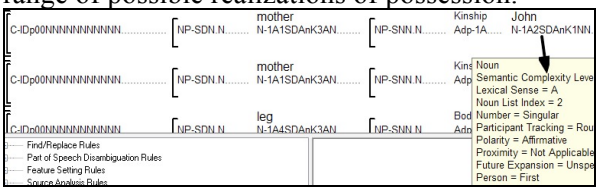
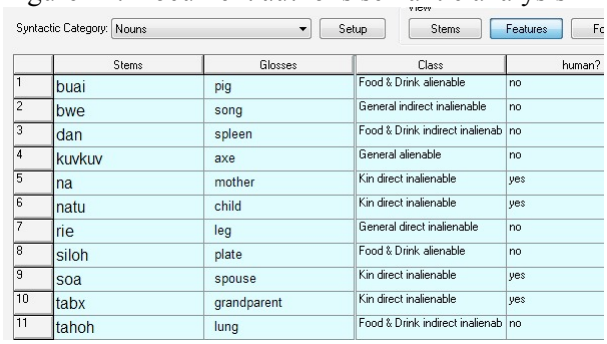


Figure 11: Document author's semantic analysis



	Stems	Glosses	Class	human?
1	buai	pig	Food & Drink alienable	no
2	bwe	song	General indirect inalienable	no
3	dan	spleen	Food & Drink indirect inalienab	no
4	kuvkuv	axe	General alienable	no
5	na	mother	Kin direct inalienable	yes
6	natu	child	Kin direct inalienable	yes
7	rie	leg	General direct inalienable	no
8	siloh	plate	Food & Drink alienable	no
9	soa	spouse	Kin direct inalienable	yes
10	tabx	grandparent	Kin direct inalienable	yes
11	tahoh	lung	Food & Drink indirect inalienab	no

Figure 12: Maskelynes noun classes in lexicon

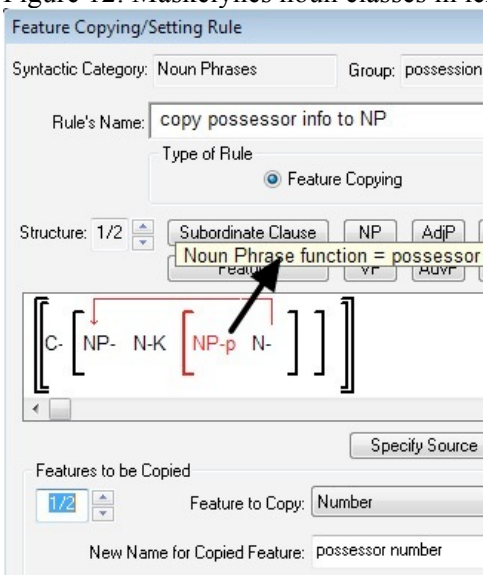


Figure 13: Feature copying rule

The preparation of features needed by the surface rules often has a final source: feature copying rules. In this case, the surface rules for choosing the correct possessor suffix to add to the main root for directly possessed nouns (section 1 of Figure 9) must have access to the person and number of the possessor. Figure 13 shows a rule that copies the number of the embedded possessor to the NP level of the main phrase. Note that the resulting feature will be called "possessor number." A similar rule copies the person. Figure 13 also shows (in the mouse-over tooltip pointed to by the arrow) an addi-

tional Noun Phrase target feature not mentioned yet: the "Noun Phrase Function." This feature can have various values in Maskelynes, but will be set to the value "possessor" for these examples by a rule that we will not discuss (which is, in fact, the main point: we can define these features and use them without worrying about the rules that will correctly set their values). The fact that we did not mention (or even think of!) this feature earlier points out what must be obvious: the need for target features often only becomes evident as you consider how to build an easy surface rule, or even as you think about how to write the rules to set other target features (in section 4.4 below). In all, there are three sources of features to be used in the surface rules: lexical features, target features defined by the user and set by rules, and features that were copied from other constituents using feature copying rules.

4.3 Writing Surface Rules

Space prevents us from presenting all of the surface rules; we concentrate on the rule that produces the possessive suffix for directly possessed nouns (section 1 of Figure 9). Figure 14 shows the surface table rule that adds possessive suffixes to directly possessed nouns. The three arrows point to mouse-over pop-ups that detail which feature values are referenced in the indicated row or column. The top-left corner cell of the table is a catch-all cell; the input must match this corner cell or the rule will not apply. In this case, the requirement is that the "Noun Possession Type" feature of the noun must be "possessed." The "Noun Possession Type" is the target feature we discussed above. Later, we will need to write a rule to set its value to "possessed" for directly possessed head nouns. But the main point here is that this particular surface rule that adds the suffixes is extremely simple if we assume the presence of that feature. The pop-up that appears on a mouse-over of the first row shows that this row refers to the "possessor person" feature on the Noun Phrase. This feature (which is a copy of the person of the possessor noun) was created above using the feature copying rules. Likewise the first column refers to the "possessor number" feature on the Noun Phrase, also copied using the feature copying rule described above. The table rule simply defines the correct suffixes for each combination of Person and Number.

A rule of similar complexity will add the genitive proclitic (as for *hX-vanuan*) for the case when the Noun Possession Type feature is "genitive proclitic." A separate rule will add the possession word (for example, *hagw or sagw)* when the Possession Type is "none." In general, we have identified the three major realization cases, created a target feature to reflect these choices, and then wrote three different surface output rules to actually realize each choice. Each rule is relatively simple (once a proficiency in using LA is attained). At this stage we have not even worried about the rules that set the target feature that makes these surface output rules possible.

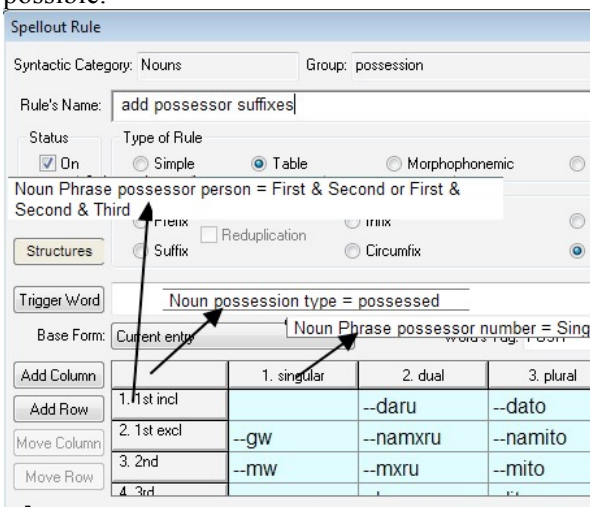


Figure 14: Table rule for possessive suffixes

4.4 Writing Rules that Set Target Features

The surface rules described above are simple because a well thought-out system of target features is assumed. The target features can come from the lexicon, from user-defined target features, or from feature copying rules. Those that are user-defined need to have their values set with rules. The "Noun Possession Type" feature is set by rules (Figures 15-18) that examine the class of noun and the realization type (pronoun or full noun) of the possessor.[4]

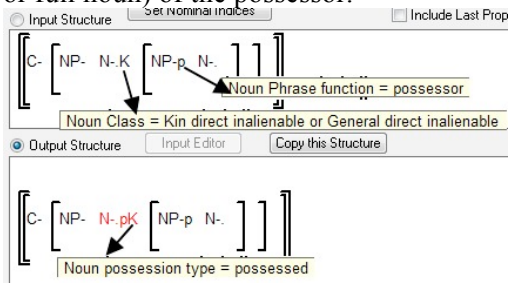


Figure 15: direct possession

Figure 15 corresponds to the directly possessed case in 1A of Figure 9. Note the "Noun

[4] For the workshop presentation, we will go over the Noun Possession Type rules in greater depth and/or use a completely different example related to Asian linguistics.

Possession Type" is attached to the head noun. Figures 16 and 17 correspond to the indirectly possessed cases in Figure 9 in which possessive suffixes are used; note that the output feature is attached to the possessor noun. And finally Figure 18 corresponds to the indirect case where a genitive proclitic is used.

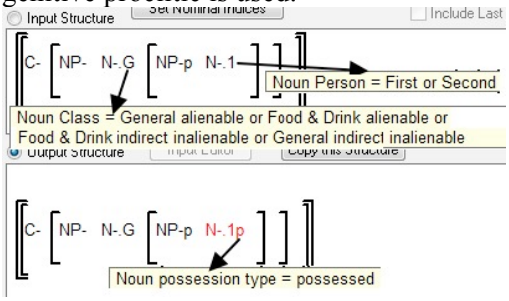


Figure 16: Indirect with possessive suffix on possessor (1st or 2nd person possessor)

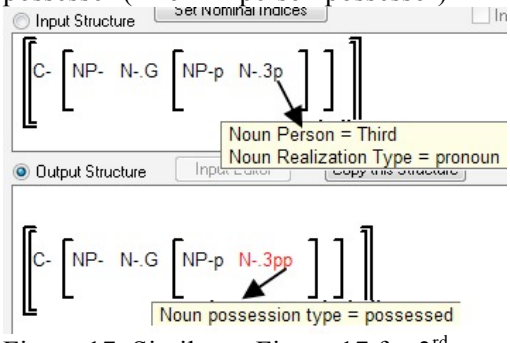


Figure 17: Similar to Figure 17 for 3rd pronoun

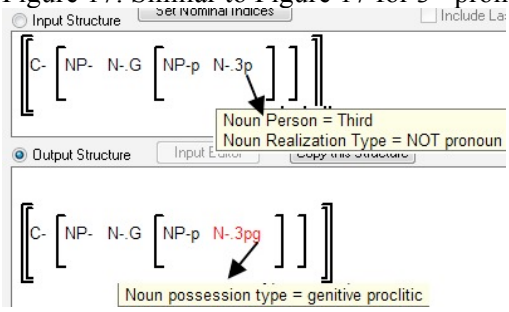


Figure 18: Indirect with genitive proclitic

We will not pretend that these rules are easy to follow in a paper with such limited space. We will be able to describe the rules better at the presentation. The main point is that the rules themselves are relatively simple to construct once a certain level of familiarity with LA is attained. The progression we presented here is typical. The surface rules (section 4.3) are easy because we assume the presence of well thought-out target features. The rules that set these target features (section 4.4) can be simple as well. Thus the entire process is often straightforward once the methodology is learned. The workshop presentation will focus on using LA to implement this methodology.

4.5 Testing the Acquired Knowledge

The final stage of knowledge acquisition is to test the system by generating target text from the example sentences that were previously authored (section 4.1). Figure 19 shows the results of generating from the authored examples. The top text box contains the target translations, which in this case match the expectations from Figure 9. The yellowish mouse-over pop-up that appears when the cursor is placed on *a-na-gw* shows all of the rules that were applied that were related to that target word. LA also includes a breakpoint mechanism that allows the user to step through the application of a rule. These debugging tools (Allman, 2010) are invaluable in the knowledge acquisition, test, revision loop that is typical.

Conclusion

This paper along with Beale (submitted; 2011) describe the two main modes of language description in LA: 1) using the provided semantic elicitation corpus to guide knowledge acquisition, and 2) using naturally occurring texts that exemplify interesting target-language linguistic phenomena. We demonstrated the latter process in this paper using possession in Maskelynes as an example. In future publications we intend to present other linguistic phenomena, summaries of LA's overall use to document a specific language, and further tutorials and descriptions of LA itself.

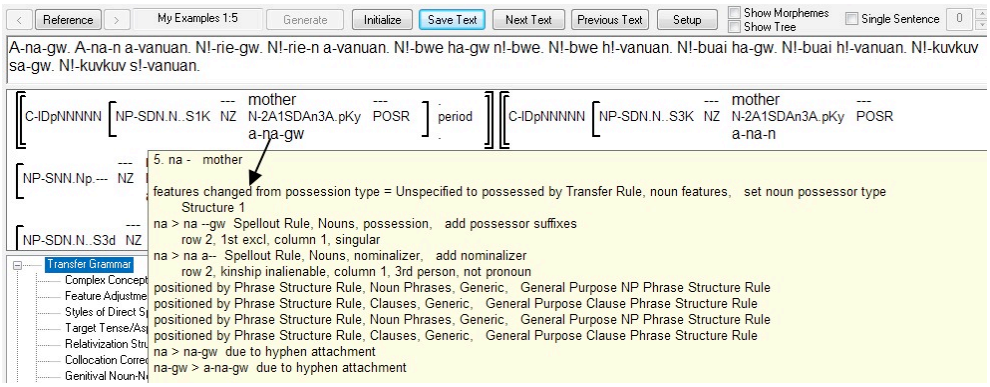


Figure 19: Output translations

References

Tod Allman. 2010. The Translator‘s assistant: a multilingual natural language generator based on linguistic universals, typologies, and primitives. Arlington, TX: University of Texas dissertation.

Tod Allman and Stephen Beale. 2006. “A natural language generator for minority languages,” in Proceedings of SALTMIL, Genoa, Italy.

Tod Allman and Stephen Beale. 2004. “An environment for quick ramp-up multi-lingual authoring,” International Journal of Translation 16(1).

Stephen Beale. Submitted. “Documenting endangered languages with linguist’s assistant.” Language Documentation and Conservation Journal. Draft available at: http://ilit.umbc.edu/sbeale/LA/papers/DEL-for-LDC-journal.pdf

Stephen Beale. 2011. Using Linguist’s Assistant for language description and translation. In Proceedings of The 5th International Joint Conference on Natural Language Processing (IJCNLP-11) Demonstrations, Chiang Mai, Thailand.

Stephen Beale, S. Nirenburg, M. McShane, and Tod Allman. 2005. “Document authoring the Bible for minority language translation,” in Proceedings of MT-Summit, Phuket, Thailand.

Emily Bender, S. Drellishak, A. Fokkens, M. Goodman, D. Mills, L. Poulson, and S. Saleem. 2010. “Grammar prototyping and testing with the LinGO grammar matrix customization system,” in Proceedings of the ACL 2010 System Demonstrations.

Sheryl Black and Andrew Black. 2009. “PAWS: parser and writer for syntax: drafting syntactic grammars in the third wave,” http://www.sil.org/silepubs/PUBS/51432/SILForum2009-002.pdf.

Tomoko Ishizuka. 2010. Alienable-Inalienable asymmetry in Japanese and Korean possession. *University of Pennsylvania Working Papers in Linguistics.* Volume 16: issue 1.

Frantisek Licthenberk. 1985. “Possessive constructions in Oceanic languages and Proto-Oceanic,” in Austronesean linguistics at the 15th Pacific Science Congress, Canberra: Pacific Linguistics C-88.

Marjorie McShane, Sergei Nirenburg, Jim Cowie, and Ron Zacharski. 2002. “Embedding knowledge elicitation and MT systems within a single architecture,” Machine Translation 17(4), pp.271-305.

Katharina Probst, Lori Levin, Erik Petersen, Alon Lavie and Jaime Carbonell. 2003. “MT for minority languages using elicitation-based learning of syntactic transfer rules,” Machine Translation 17(4), pp.245-270.