Supporting sign language in Human Computer Interaction,

Understanding the challenge

Abstract

Sign language is at present little used in human computer interaction although it would be of great value to do so for those speaking sign language. Seeing the great value implementing sign language in human computer interaction would have for sign speakers, doing so seems a worthy endeavor. The question that arises once this becomes the intent is what difficulties one would face doing so. This thesis investigates what fundamental differences there are between sign language and natural language in the fields of morphology phonology and syntax. Having made an analysis of these differences it then proceeds to investigate what challenges these differences would pose for trying to implement human computer interaction using sign language from both a receptive and a productive view.

Table of Contents

Introduction

In recent decades computer use has grown exponentially. The computer has grown from something used at universities for specific research purposes to a household device that almost everyone has in their house, one that is essential for a great variety of everyday things. The computer has grown swiftly into everyday life for most people, however for people with handicaps the computer its developments are less natural. For those who suffer handicaps such as partial paralysis solutions are often found. It is clear they cannot use a normal computer normally and alternatives are developed for these people.

However not everyone who is different from the norm has such an obvious handicap when using the computer. A specific group that has drawn my interest are those people who are dependent on sign language. Sign language is in no way inferior to normal language when it comes to powers of expression or relaying information.

However it is clearly different from normal spoken and written language in many ways, syntactically, morphology, phonology.[1] Sign language is the native language for many deaf people, a language they are far more adept and comfortable using then spoken language. However computers rarely offer any form of support to sign language users. There has been some research into automatic recognition of sign language but that is where it ends. Given the large amounts of signers, the growing importance of the computer and the advantages of speaking in your native tongue eventually absorbing sign language into the computer to allow signers to speak in their native tongue will be necessary.

Sign language is clearly different from spoken languages. What fundamental problems the differences in structure of sign language would present in human computer interaction is largely unknown. Given the desire to allow sign language to be used in human computer interaction, getting a good view of what fundamental challenges would arise here seems necessary. Given the enormity of any language this is not a question that can be fully answered easily if ever. However a clear part of the fundamental difficulties, especially in some fields, can certainly be found. As such I'll seek to answer the following research question;

What fundamental challenges would one face when trying to support the use of sign language in human computer interaction, looking at structure differences between spoken and sign language within the fields of phonology, morphology and syntax?

To answer this question, first we have to be aware of an important split in the question. There are two large areas here that need to be looked at. The computer has to be able to interpret the signs provided to it by a human signer. Furthermore the computer has to create signs that the user can interpret. These are the receptive and the productive parts of the interaction. In other words, a computer is a sender and a receiver of messages. These two tasks of the system can be viewed separately, as long as the context in which they operate is the same.[2][3] I will set out to describe the known structure differences in the fields of phonology morphology and syntax. With these differences I will seek to find out what fundamental challenges these structure differences imply for both receptive and productive human computer interaction. What fundamental problems arise from the differences in structure when trying to produce a sign, and when trying to interpret a sign. Then, combining the conclusions for receptive and perceptive human computer interaction, a clear overview will be provided of the fundamental challenges faced, in regards to the fields mentioned earlier, when implementing sign language in human computer interaction.

Structure differences in phonology between sign language and spoken language.

It is obvious that the phonology of sign language is not, as spoken language, based on the production of sounds as the base unit. However the structure of the language need not necessarily be different based on that. However spoken language is linear, it can be broken down into sequential sub morphemic segments. Sign language phonology begins differing from spoken language here. Words within sign language are made up from base aspects, like spoken language uses sound. However the aspects that make up signs in sign language are executed at the same time. Instead of one assembly of possible sounds that follow each other up. A part of a sign cannot be viewed separate from the sign as a whole.
However a, purely imaginary, division can be made. Each sign is composed of several actions that, together, make the sign. Actions, that are the dez (Designator, what acts), sig (signation, the actual action), and tab (The tabula, the place of the sign). An essential thing to keep in mind here is that facial expression, eye gaze and head action are just as much subject to these things as hand movement. This means that if I make a sign for 'you angry' the angry (action) has to be executed by the designator (who acts) that the you referred to. It is one sign, and thus the actions that compose it are the same over the whole sign.
Things such as facial expressions are grammaticalized within sign language and refer to each of these actions just as hand movement does. So if my face signs the angry, and a hand sign indicates you then the angry and the you are connected. The angry has something to do with the you.

Next to this separation into what (imaginary) actions are present in each sign there are also three distinct elements of each hand sign. Namely the hand shape, location, and movement. These elements together form a hand sign, where the combination of the three at one moment in time (not linearly following each other) indicates the meaning of the sign. Quite different from spoken languages that are arranged linearly in the dimension of time, sign language is more then just linear and several aspects of a sign have to be interpreted at the same time to get its full meaning. To be able to work with this a sign, at a moment in time, needs to be viewed as a unitary act, an act that can mean more then any single word.

The fact that several, in themselves lacking the full meaning, elements of a sign make up a meaningful sign and that a change in any of these elements can cause a change of the sign tells us something else though. It shows us that the principle of duality of patterning exists in sign language as well. This means that the elements together make up a more meaningful signs. The whole is more then the sum of its parts. Furthermore it means that constraints apply to these smaller elements. Like spoken languages have constraints on what sounds can follow each other, sign language has constraints on what sign elements can be used consecutively, or at the same time.[4][5]

So what structure **differences** do we see in sign language, compared to spoken language, when looking at phonology?

1. The base elements of a sign are not sequential sub morphemic segments. They cannot be broken down into sequential parts. Instead several, independent, elements have to be viewed at the same time. Basically, the use of the 3 dimensions available in hand movement.
2. The base units of a sign, happening at the same time, are all interconnected. One cannot have one element of a sign refer to one actor and another part of the same sign to another.

Structure differences in morphology between sign language and spoken language.

Sign languages were once believed to be simple, iconic, languages. Languages with little to no morphology or real complexity unable of advanced grammatical structure where each part of a sign could be viewed separately without loosing meaning. This was half a century ago, and since then science has come a long way. It is clear now that sign language does have a clear grammatical structure, and a high level of morphological complexity.

The important thing to remember from this is that insights change. In the field of morphology of sign language the current theory is based on solid research. At the same time all this modern research makes it clear in the papers that the study of the morphology of sign language is still in its infancy. The current insights might well be proven wrong in the future, sign language morphology might prove to work differently then currently believed.

A thesis like the current one is hardly the place to actually study the morphology of sign language in depth. However it is important to keep in mind that the conclusions drawn here are based on the present view of morphology of sign language which (according to its researchers) is still incomplete and in its infancy.

So what is the present state of the morphology of sign language?

At present it is believed that sign language has a very complex grammatical structure. The formational characteristics of a sign language are a difficult interweaving of rules and there is no simple way to describe it. No simple semi lattice, which has the property of continuing, non intersecting, divergence can represent it. See figure 1 for but a small part (and already slightly outdated) of the formational characteristics, it does however give some idea of the difficulties; (Fig 1 taken from [4])

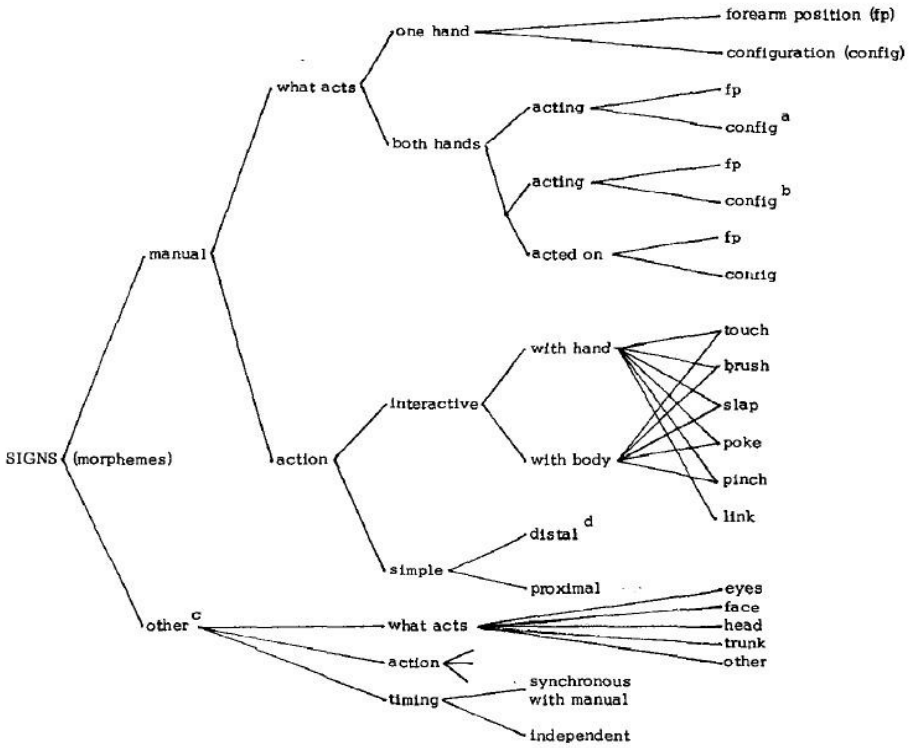

Figure 1 **Formational division of sign language morphemes.**

If the diagram in figure 1 were extended further to the right one would see that many of the present lines would come together again. A sign action that is pointing to a part of the body brings sign action and body together again, and a hand has to execute it. The diagram in figure 1 is by no means complete (just think of the many possibilities that can fall under the node 'other'). However it is hopefully illustration enough for the great complexity of the morphology of sign language.

The present view in sign language morphology is that, complex as it is, it does greatly resemble some spoken languages. Sign language morphology is far more complex then say, English or Dutch morphology. However when looking at languages as Navaho, Swahili or Hebrew sign language shows great similarities in the complexity, and form, of morphology. The list of morphemes with which sign language works has a similar finite form determined by the lexicon and grammatical rules of the language.

In short, sign language has a complex morphology but it is not significantly more complex or different from the morphology of some spoken languages, such as Hebrew or Navaho. There are no fundamental structural differences known between sign language and spoken languages in the field of morphology. [4][5][6]

Structure differences in syntax between sign language and spoken language.

Syntax lies at the basis of much of what makes human language human language, the possibility to make a limited number of base utterances into an unlimited amount of sentences, of course limited by what one feels is common sense and what is utter nonsense. This possibility comes from recursiveness in human language, a feature which I will not describe in great detail here but on which many papers are written and extensive information can be found.

Recursion appears in two basic forms in human language, coordination and subordination, and for the longest time (and still) both forms are considered essential to a full fletched human language and for the power of expression this means.
Hence my surprise was quite profound when I found the first research papers that said that sign language actually lacked the mechanisms for subordination and recursiveness. The research was quite dated, and modern research clearly waylays it. The syntax of sign language was so very different that ,looking at sign language like a spoken language, the options for recursiveness were not recognized when they were there, in different form. This did however spark my interest, if the language was so different there might be some clear differences in syntax.[5][7][9]

Sign languages structure and rules are quite different from that of spoken languages. What we are used to is that changes in word order or special words make the syntax clear. You can look at the word order or formative words and see what sentence follows, or is subordinate, to another.
Research into sign language has shown this is not the case there. Sign language uses similar hand signs for both subordinate and coordinate sentences, furthermore there are no clear hand signs distinguishing a transition from one to the other. Sign language identifies these differences not by specific words or changes in hand signs, instead it identifies such syntactical structures by differences in expression or head position.

What this means, aside from the fact that more then just the hand signs are important, is that the non-linearity of sign language is reflected here as well. A sentence can both be made and at the same time its relation to other sentences can be displayed. A sort of trigger moment (departing from normal head / expression) is turned on to make clear the syntactical position of the clause. Sign language uses, non manual, marker moments in its grammar for recursiveness.

The specifics of how this works are far better illustrated in the words of real linguists then my own;

"Liddell identified the non-manual marker of relative clauses as consisting of raised brows, a backward head tilt, and a tensed upper lip. This non-manual marker co-occurs with the material of the relative clause. In addition, Liddell argued that relative clauses in ASL are not like those of English

In particular, he showed that the noun that is modified by the relative clause occurs inside the clause in ASL ... rather than outside it as in English."[5]

Aside from this difference with spoken languages, where word order or specific words indicate the syntax, sign language has some more very clear differences from Indo European languages in terms of syntax.
The primary of these is the fact that sign language has no clear cut distinction between words and sentences, a word can be a sentence in itself, and there is no set of words (verbs, noun's etc) that have to be in a sentence. Instead actions, an actor, and such have to be in a sentence (but these can at times be presented in a single sign).
Although differences in syntax like this from English are quite profound there are other spoken languages that do have similar structures, one can think of Navaho or Hebrew as examples of spoken languages with similar structures. As such this clear difference in structure from Indo-European languages will not be taken along. This research question goes into the differences between spoken languages and sign language. When the sign language compares to some, but not all, spoken languages then a fundamental difference and related challenges for human computer interaction are not likely to be found. No challenges at least that are not already faced by programs that support languages such as Hebrew. Seeing mainstream programs such as windows support such languages makes it highly unlikely there are unknown challenges related to these languages..[5][6][8]

A final important aspect where sign language structure is different is prosody. Prosody involves both syntax and phonology. In spoken languages prosody is the rhythm and intonation in sentences. It allows speakers to emphasize a word, or part of a sentence, and to split a sentence into multiple parts. Aside from such functions it also tells us a lot about the context of a spoken sentence. Is the speaker angry, Sad, Happy?
Research has shown that sign language has something quite similar to prosody. Sign language cannot work with a change in intonation, nor with an increase in volume. Sign languages systems works with facial expressions, body posture, and rhythm in signing to achieve the benefits of prosody.

What is very important to realize here is that sign language facial expressions are not the same as a person just looking angry. The facial intonation and body language used in sign language are grammaticalized. They are fixed, systematic, and bound by clear rules. [5]

So what structure **differences** do we see in sign language, compared to normal language, when looking at syntax?

1. Sign language syntax does not rely on word order. Instead facial expressions indicate the syntactical position of part of a sentence.
2. Sign language syntax is not linear, instead it is multi processed. A sign can be formed and its syntax formed at the same time.
3. Prosody in sign language is again part of facial expressions, and not in the hand signs themselves. Furthermore it is heavily grammaticalized, in contrast to spoken languages which are a bit more lenient in rules about voice intonation, pitch, and such.

Fundamental differences in structure between sign language and spoken language looking at the fields phonology, morphology and syntax

Having looked at the 3 fields above separately to try and discern the fundamental differences with spoken language we now have the ability to make one encompassing view of what these differences are. The explanation of where these differences come from should be clear by now, as such there will simply be a list of fundamental differences here.

1. The base elements of a sign are not sequential sub morphemic segments. They cannot be broken down into sequential parts. Instead several, independent, elements have to be viewed at the same time. Basically, the use of the 3 dimensions available in hand movement.
2. The base units of a sign, happening at the same time, are all interconnected. One cannot have one element of a sign refer to one actor and another part of the same sign to another.
3. Sign language syntax does not rely on word order. Instead facial expressions indicate the syntactical position of part of a sentence.
4. Sign language syntax is not linear, instead it is multi processed. A sentence can be formed and its syntax formed at the same time.
5. Prosody in sign language is again part of facial expressions, and not in the hand signs themselves. Furthermore it is heavily grammaticalized, in contrast to spoken languages which are a bit more lenient in rules about voice intonation, pitch, and such.

Now that we have these 5 structure differences the question arises, what fundamental challenges each of these differences might present for human computer interaction.
As mentioned before we have to look at what challenges each of these points might imply from both a receptive and productive human computer interaction viewpoint. I will seek to answer what challenges might be faced receptively first, then I will look at what fundamental challenges there will be when trying to generate signs with the computer.

Fundamental challenges for sign language in human computer interaction, a receptive view.

Having assembled a list of fundamental differences between spoken language and sign language we will be taking a view at what challenges these differences actually mean. I will view each of the 5 found differences in order and explain what challenges, or lack thereof, they might present from a receptive view.

To start with the first;

The base elements of a sign are not sequential sub morphemic segments. They cannot be broken down into sequential parts. Instead several, independent, elements have to be viewed at the same time. Basically, the use of the 3 dimensions available in hand movement.

When it comes to recognizing parts of sign language interaction, like solely facial expressions, or only hand signs, no great obstacles are met. Of course recognizing these is not easy, and there is a clear margin of error. However when applying Hidden Markov model (HMM)-based recognition good results are achieved. With hand sign recognition getting correct results above 90% even without clear adherence to grammar constraints [12] no real worries exist that this may not be done successfully.
Other methods (such as dynamic time warping) that require linear representation cannot be used. However seeing the success of Hidden Markov models in recognizing hand signs, considerably more successful at recognizing these then at recognizing spoken language [13], there is no reason for worries.

Facial recognition is not quite as successful as hand recognition for sign language but it has a high enough success rate to be useful for practical purposes. Furthermore there is still considerable room for improvement. [5]
The use of multiple dimensions, or the input itself, does not seem to be the problem however.
The real issues arise when one is forced to combine multiple of these recognition models, needing them to recognize signs at the same time. We have seen that parts of the sign cannot be viewed independently and have their whole meaning registered. For a long time it was believed to be very complex, close to impossible, to view all these parts at once. However with recent gains in processing power as well as further research into multi event processing it has become clear that it is possible to do sign recognition on multiple threads correlated solely by their time aspect. Further links between the multiple event threads only make the recognition more accurate. [12]

Although these methods have proven fairly successful, and are likely the best way to proceed for anyone wanting to make an actual implementation of sign based human computer interaction, some important notes about them remain.
The method described in [12] works similar to an interval algebra network. Interval algebra networks work well within limited scenarios. However for general interval algebra networks finding both the scenario and the feasible relations has been shown to be NP-complete. [14][27] Therefore, any implementation here may face great difficulty.
The current methods to process sign language without abstracting from either facial or hand gestures, or the creativity of humans, have the challenge of NP-completeness. Although language sentences may not gain such length that it is not computationally feasible to work with them in such a way having your solution be NP-complete is never comfortable.
Anyone wishing to make a sign language recognition system would need to be well aware of this particular problem, and design their system to be as efficient as possible in handling this.[11][12][27]

The second difference is only slightly different from the first. It is an added constraint between the several signing parts, such as between face and hands.

The base units of a sign, happening at the same time, are all interconnected. One cannot have one element of a sign refer to one actor and another part of the same sign to another.

Fortunately, unlike the previous difference which brought great difficulties in the form of NP-complexity, this difference between sign language and spoken language structure poses a wholly different kind of issue. As we have seen in [12] [14] the present view on multi-event handling at a same, related, temporal dimension leads to an NP-complete challenge.
However it has also been shown that further similarities, aside from just the temporal one, can give a higher success rate in finding the relation between two events (such as two signs) at a computationally lighter cost. [14]
Although it does not solve the issues we found before it does provide a chance to make better predictions of what is being signed as well as being able to handle a slightly larger number of relations.
A challenge will lie in finding a good way of exploiting these possibilities, something on which I could find no relevant literature. All the more reason for anyone seeking to implement sign language for full fledged human computer interaction to take a look at the opportunities this presents.[12][14]

The third difference that was found is considerably different from the first two, stemming from the field of syntax instead of from phonology.

Sign language syntax does not rely on word order. Instead facial expressions indicate the syntactical position of part of a sentence.

To answer this question we will start by looking at what the syntax of normal language entails in the receptive side of its human computer interaction.

Within current speech recognition syntax is used to tie sentences together. It helps a computer to know when a sentence ends and a new one begins, as well as what the relation is between the sentences. Furthermore it gives an indication of when a speaker is done and another speaker can start speaking. The way syntax, of spoken languages is used for this is by looking at the combinations of words, or words themselves, and likelihood that these will be followed by a certain syntax rule, such as a new sentence.

Clearly looking at words and deriving the syntax from these (such as is commonly done with spoken languages) will be impossible. The sign language syntax specifically does not rely on word order. As such the current methods of syntax deduction that are used in speech recognition will not work for sign language.
However research has shown that actual triggers that indicate syntax would work far more effectively then the current statistical methods, especially when trying to deduce syntax over multiple sentences.[15] And this is how sign language syntax works. When trying to implement sign language in human computer interaction the syntactical difference between sign language and spoken language is such that the presently used methods for recognizing syntax cannot be used. However the way syntax is done within sign language is something computers can handle well. The challenge lies in developing a new system aimed specifically at sign language syntax, fortunately doing so should not be overly difficult. [15][16]

The fourth difference we saw was once again related to syntax, namely the following.

Sign language syntax is not linear, instead it is multi processed. A sentence can be formed and its syntax formed at the same time.

This specific difference has two sides to it. Firstly it is once more a multi process event thread. Difficulties pertaining to this, and the complexity of this problem, have been described at the first discussed difference of this chapter regarding sequentiality of the parts of a word and in [12]. As such I will not go into detail in regards to this issue. A similar challenge of NP-complete complexity as described before will present itself due to this difference in syntax.
Aside from this challenge the question remains if this difference in syntax causes other challenges. As we can see much of this difference lies close together with the fact that not words, but facial expressions lie at the root of sign language syntax, seen in our previous point. [15][16]
The commonplace statistical analysis do not work, although we see other methods are available, and the complexity is NP-complete. Aside from these things no new issues arise. So earlier noted challenges apply here as well but no new ones arise.

After the issues with syntax we now come to the difference in prosody between signed languages and spoken language, namely

Prosody in sign language is again part of facial expressions, and not in the hand signs themselves. Furthermore it is heavily grammaticalized, in contrast to spoken languages which are a bit more lenient in rules about voice intonation, pitch, and such.

One of the great limitations of prosody in spoken languages, when it comes to speech recognition and human computer interaction, is the fact that they have so far been unable to get truly good results in recognizing human emotions.
Prosody can provide good information about the activity level of a user, however the emotional state itself is still out of reach. They are unable to accurately detect real time whether someone is very joyful or very angry, merely that they are very active.[17]
When it comes to prosody in sign language, and due to the high grammaticalization of prosody in the language, such distinctions can already be made for sign language.[5]

The reason this is worthwhile is because being able to recognize emotions so well is a thing that is desired in human computer interaction, and that has a good deal of useful possible applications.[18]

The present methods used for prosody recognition are based on very different factors from those in sign language, voice pitch, voice rhythm, that these will not be useful.
New methods will have to be developed to detect the precise prosody of sign language. However the primary challenge will not lie in developing the prosody recognition system for sign language.[5] Instead the primary challenge will probably be to use the options sign language gives us here, and developing applications that will be able to properly use the information about the emotional state of the user.

Fundamental challenges for sign language in human computer interaction, a productive view.

Aside from being able to receive signs the computer will also have to produce signs to be able to communicate with people in sign language. What challenges await us in this regard, looking at the differences between signed and spoken language?

Like before we will move down the list of found differences. Many of the issues found in a receptive view may also apply in a productive view. Cases such as multi-event handling are likely to be applicable both in the receptive and the productive view. In those instances it will be mentioned. However the depth of the challenge this poses, if it is no different from the receptive view, will not be considered. Obviously in these cases the challenges as viewed in the receptive view will give a clear explanation of what these challenges are and entail.

Now, on to the first difference;

The base elements of a sign are not sequential sub morphemic segments. They cannot be broken down into sequential parts. Instead several, independent, elements have to be viewed at the same time. Basically, the use of the 3 dimensions available in hand movement.

The first thing to note here is that the earlier noted issues of the receptive view, in regards to complexity, need not necessarily apply.
One can create sign language like its received from human speakers. The facial signs and hand signs happen at the same time but are recorded independently. The parts of the sign cannot be viewed independently from each other. The temporal relation between the two are complex, it matters whether a facial expression starts before, after, or during a hand sign and the next hand sign. The complexity if one wishes to have all the possibilities and grammatically correct sentence structures to be constructible is NP-complete.[12] [27]

However one can choose to limit the complexity. This is done by applying restrictions upon the sentences that are allowed. Limitations need not necessarily be a bad thing in this. A computer should not want to make very complex temporal structures and if one limits the structures the computer makes to several simple temporal combinations, the complexity can be brought down to one with polynomial processing time.[27]
Another choice arises, in present day techniques of natural language generation the computer will generate its communication based on a corpus, a dictionary of

possibilities from which it draws. In the case of the corpus of sign language a choice has to be made.
The corpus can be composed of a relatively small number of units. The hand signs (words) themselves along with the associated facial expressions as separate corpus entities. The clear rules of grammar would have to be laid out, to ensure the system knows which signs can happen at the same time as which and what these mean together; this seems possible. However this means more computation at a user level.
The other option is to have a huge corpus, existing of all the combinations of signs with the possible facial expressions they can have. This is obviously a huge corpus to build, and it is probable one would have to trim down the broad possibilities of the language to make this corpus at all possible. Furthermore great difficulty will likely arise in anticipating all the possibilities. However that huge corpus es of sign language can be made has been shown, a good example of this in practice is; http://webmail.gebarencentrum.nl/woordenboek.asp?color=y&idrelease=28.
Does one choose to construct a huge (and memory intensive) corpus that will lead to easier programming and less processing required for the end user, or go with more computationally intensive but also smarter creation algorithms that actually do manage (part of) the temporal aspect ?[12][19][20][26]

The second difference in structure remains largely similar to the first. However, the question is what challenges it will lead to with the current viewpoint. What are the challenges posed in producing signs by the fact that;

The base units of a sign, happening at the same time, are all interconnected. One cannot have one element of a sign refer to one actor and another part of the same sign to another.

What this means is that the multiple parts of a sign that make up the sign have the same context, if one chose to have the computer use the language structure as human signers use it. As we saw before the computer need not necessarily follow that grammatical structure to be able to produce signs. If one chooses to, as in the example given before, create a huge corpus where whole signs (not the separate parts) are saved then the earlier difference in structure poses no direct challenges. However, one would have to closely keep this rule in mind and ensure the contents of the created corpus follow it. [19][20]
If one chooses to have their computer generate the signs based on different parts of the sign, as people do, more challenges crop up.
Although the use of meta data and context by computers has grown tremendously over the last decades, as well as their successes, these systems are

far from perfect. Especially when a computer seeks to have a dialog with its users using context and surroundings difficulties still exist in abundance.

The computer is certainly growing better at handling these kinds of difficulties but there is still a lot of work to do. Furthermore there is little present research into what context and meta-data do when presented with the challenge of two separate 'parts' of a whole. Neither part is useful without the other, and so the data only has a relation when the two are taken together.
One has to choose here. One can go for a large corpus of signs, where parts of a sign such as hand sign and facial expression are saved together in all possible combinations, causing a large corpus. The choice can also be made to research what consequences handling context in a different way would have on the techniques used in speech recognition. What parts of it can be 'recycled' to fit sign language and what would have to be made, and make those parts that need to be.[21][22][23]

Now that we have looked at the phonology related differences it is time to move on to the ones related to syntax, namely;

Sign language syntax does not rely on word order. Instead facial expressions indicate the syntactical position of part of a sentence.

Just as the sign language syntax poses no great difficulties from a receptive point of view it poses little difficulty from a productive point of view. Although most present day syntax systems are adapted towards sentences were the words themselves carry the syntax this is rarely how the computer actually creates and handles the syntax. Often enough when a computer is producing language, and related syntax, it adds in formalized rules (that are not shown to users) to act as syntax. From a computer point of view, this is a simpler solution to an otherwise difficult problem. A few extra characters added in between the words is no real difficulty to the computer whatsoever when it processes them linearly. Many computer solutions for the handling of syntax already separate syntax and the words within the computer, much as sign language does, and then proceed to merge them to get the desired text results out.
Although any technique would have to be adapted for both the syntax and specific grammatical rules of sign language the underlying technique should pose no real challenges. Some more recent syntax parsers actually use underlying methods quite similar to sign language already, and there should be no great fundamental challenges to adapt such parsers. [15][24]

Having viewed this difference from a productive view we move on to;

Sign language syntax is not linear, instead it is multi processed. A sentence can be formed and its syntax formed at the same time.

Looking at this difference, much like with the last difference, the computer will not work with sign language, or its syntax, quite as humans do. However when it comes to creating sentences, if clear grammatical rules are specified, no large problems or challenges are expected. The computer can produce the syntax 'around' the language using clever implementation of tree based designs as long as the grammar rules are clearly specified. Obviously it will be a great challenge to specify grammar rules to such an extent. However, it is certainly not impossible nor any different from the challenge for any spoken language attempting this. [4][5]
There will need to be clear adaptations to possible solutions to work better with sign language , but there should be no great technical inhibitions for a computer to create sign language based on this.

Obviously, the great amount of problems that spoken languages face are not suddenly gone for sign language.

Having seen the last two syntax related differences give no real fundamental challenges we now look at the last difference from a receptive view, the one related to prosody.

Prosody in sign language is again part of facial expressions, and not in the hand signs themselves. Furthermore it is heavily grammaticalized, in contrast to spoken languages which are a bit more lenient in rules about voice intonation, pitch, and such.

We already saw that parts of the language being in the facial expressions need not pose any challenges if one is willing to handle the language formation in a computerized way, which is different from how humans are currently believed to do it.
One of the things this would entail was that the computer needed a very clear, formalized, grammar structure for it to follow. Strict rules that it could adhere to that would give it the ability of creating at least acceptable language. This last demand seems to be partially met by the final structure difference. Obviously a clear sacrifice has to be made here. The computer will not be generating the language in the way we, humans, do. Whether this will make it totally impossible for the computer to generate good dialog is something that we cannot answer here, although personally I suspect it will not even if the dialog might be differently constructed.
The challenges that apply to spoken language generation remain as well. In short this structural difference does not, at present, seem to be the cause of any great fundamental challenges. An important thing to remember however is that as natural language processing progresses so may the difficulties that these differences could bring.
The languages are different and totally different mathematical techniques and other methods can be used to work with them ,even if current techniques are largely the same. New techniques that are developed might not work with sign language, or spoken language. Present day techniques are adaptable. As such, at present the prosody differences do not cause any difficulties from a productive point of view though.[22][24]

Fundamental challenges when implementing sign language in human computer interaction?

Now that the challenges from both a receptive and a productive point of view have been found it is time to make an overview. As we saw quite a few challenges cropped up from different differences between sign and spoken language. Indeed some challenges were encountered more then once. Obviously such challenges need not be named 5 times in this overview.
Similarly copying down every single difference found above in a neat list here would not be useful. To fully understand the differences and challenges there it is best to just read them directly, they cannot be simply summarized in 1 sentence. What we will try to do here is to name the things at the core of the challenges described above. The core concepts that one has to be wary off, and that underly the specific challenges.

So, what *are* those cores at the heart of the challenges one would face?

1. The complexity of recognizing sign language without abstracting from possible, if rarely used, options has an NP-complete complexity.
2. Sign language has a great deal of strong grammatical rules that can be used for better automatic recognition. Techniques and methods to work this out have yet to be made though.
3. Sign language differs so significantly from spoken language in several regards, syntax primarily, that from the receptive point of view, many contemporary techniques are in need of adaptation. The concepts underlying them are largely sound but any implementations need to be heavily adapted to fit sign language, severely limiting bootstrapping possibilities.[23]
4. Much like the syntax of sign language is different so is its prosody. This allows better recognition of a users emotions and the like. Actually using this information will be a challenge as its not yet used in present day programs to any extensive extent.
5. When producing sign language ,with a computer, a choice will have to be made between limitations on the temporal and grammatical structures allowed or NP-complete complexity.
6. When producing sign language, with a computer, a choice will have to be made between more work in forming a corpus or a computationally heavier and more complicated program.

7. From a productive point of view many contemporary techniques are also in need of adaptation to work for sign language. Although the techniques work the implementations of them need to be completely overhauled. Again this severely limits bootstrapping.
8. Any advances in natural language generation and natural language interpretation would have to be viewed separately for sign language to ensure the new techniques and concepts would work for the different structure of sign language as well.

This list should sum up the primary fundamental challenges for anyone wanting to implement sign language in human computer interaction when looking at the fields of syntax, phonology and morphology from a receptive and a productive point of view.

My personal advice to anyone would be to certainly develop sign language in human computer interaction if they have the means. The challenges are there, but none of them are insurmountable. Even the complexity is likely manageable, rarely will concepts be related over a big temporal structure. Whoever undertakes such a project though, should be aware of the fundamental differences and challenges these bring as well as the fact that there are bound to be a thousand challenges awaiting that any new human computer interaction project faces.

Literature

1: Wilbur, Ronnie B.,(2006) What does the study of signed languages tell us about `language'?, Sign Language & Linguistics, Volume 9, Numbers 1-2, pp. 5-32(28)

2:Frieder Nake, Susanne Grabowski,(2001) Human-computer interaction viewed as pseudo-communication, Knowledge-Based Systems, Volume 14, Issue 8, Pages 441-447, ISSN 0950-7051, DOI: 10.1016/S0950-7051(01)00140-X.

3:de Souza, C.S. (1993) The semiotic engineering of user interface languages. *International Journal of Man- Machine Studies* 39, pp. 753–773.

4:STOKOE, W. C. (1960). Sign language structure. Silver Spring, MD: Linstok Press.

5:Sandler, W. & Lillo-Martin, D. (2000) Natural sign languages. In: *The Blackwell handbook of linguistics,* ed. M. Aronoff & J. Rees-Miller. Blackwell.

6: Kyle, J. & Woll, B. (With Pullen G., and Maddix, F). (1985). Sign language: the study of deaf people and their language. Cambridge: Cambridge university press.

7:LIDDELL, S. K. (1978). Nonmanual signals and relative clauses in American Sign Language. In P. Siple (Ed.), *Understanding language through sign language research.* (pp.59–90). New York: Academic Press.

8: Wilbur, R. (1979). American Sign Language and Sign System~ Baltimore: Univ. Park Press. 312 pp

9: Liddell, Scott K. (1980). American Sign Language Syntax. The Hague: Mouton

10:S. Morrissey and A. Way. (2005). An example-based approach to translating sign language. In *Workshop Example-Based Machine Translation (MT X–05)*, pages 109–116, Phuket, Thailand, September.

11:T. Starner and A. Pentland, (1995) "Real-Time American Sign Language Recognition From Video Using Hidden Markov Models," *Proc. Int'l Symp. Computer Vision*, Coral Gables, Fla.,. Los Alamitos, Calif.: IEEE CS Press.

12:somboon Hongeng and Ramakant Nevatia, Multi-Agent event recognition, Institute for Robotics and Intelligent systems, university of southern California, Los Angeles, Proceedings of the Eighth IEEE International Conference on Computer Vision (ICCV'01)

13: J. Glass,(2003) “A probabilistic framework for segment-based speech recognition,” *Computer, Speech, and Language*, vol. 17, no. 2-3, pp. 137–152,.

14: VAN BEEK, P. (1990.) Reasoning about qualitative temporal information. In Proceedings of the 8th National Conference of the American Association for Artificial Intelligence (Boston, Mass., Aug.). MIT Press, Cambridge, Mass., pp. 728-734.

15:Bellegarda, J. R., (2004.) Statistical Language Model Adaptation: Review and Perspectives. Speech Communication, Vol. 42, 93–108.

16Tracy Hammond and Randall Davis. LADDER: (2003.) A language to describe drawing, display, and editing in sketch recognition. *Proceedings of the 2003 Internaltional Joint Conference on Artificial Intelligence (IJCAI)*,

17:Fragopanagos, F. and Taylor, J.G.,(2005), Emotion recognition in human-computer interaction, Neural Networks, 18, pp 389-405

18:R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, (Jan. 2001.) “Emotion recognition in human-computer interaction,” *IEEE Signal Processing Mag.*, vol. 18, pp. 32–80,

19:REITER, E. AND DALE, R. (1997.) Building applied natural language generation systems. *Natural Language Engineering 3*, pp. 57–87.

20:Papamichail KN, French S. (2003.) Explaining and Justifying the Advice of a Decision Support System: a Natural Language Generation Approach. Expert Systems with Applications; 24(1):35-48.

21:Jane Greenberg, Kristina Spurgin and Abe Crystal, (2006.) , **Functionalities for automatic metadata generation applications: a survey of metadata experts’ opinions ,** *Int. J. Metadata, Semantics and Ontologies, Vol. 1, No. 1, 2006* 3 Inderscience Enterprises Ltd.

22:Joseph Polifroni, Grace Chung , Stephanie Seneff (2003.), Towards the Automatic Generation of Mixed-Initiative Dialogue Systems from Web Content , EUROSPEECH 2003 - INTERSPEECH 2003 8th European Conference on Speech Communication and Technology Geneva, Switzerland September 1-4, 2003

23:M. Davel and E. Barnard,(2003.) “Bootstrapping for language resource generation,” in *Proceedings of the 14th Symposium of the Pattern Recognition Association of South Africa*, South Africa,pp. 97–100.

24:Purver, M. & Otsuka, M. (2003) Incremental generation by incremental parsing: Tactical generation in Dynamic Syntax. In: *Proceedings of 9th EACL (European Association for Computational Linguistics) Workshop in Natural Language Generation, Budapest, April 2003,* pp. 79–86. Association for Computational Linguistics.

25:Johnston, T. & Schembri, A. (2006). Issues in the creation of a digital archive of a signed language. In L. Barwick & N. Thieberger (Eds.), Sustainable data from digital fieldwork (pp. 7-16). Sydney: Sydney University Press.

26:Diane Brentari,(2002.) Effects of language modality on word segmentaton: An experimental study of phonological factors in a sign language, The 8th Conference on Laboratory Phonology, June 27-29 2002

27:Vilain, M., and Kautz, H. (1986.) Constraint-Propagation Algorithms for Temporal Reasoning. , In Proceedings of the Fifth National Conference on Artificial Intelligence, 377–382.